



**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA**

LUIZA BAPTISTA FERNANDES

**DETECÇÃO E DIAGNÓSTICO DE VAZAMENTOS DE GÁS POR REDE NEURAL
NÃO SUPERVISIONADA EM TUBULAÇÃO DE DISTRIBUIÇÃO**

**DETECTION AND DIAGNOSIS OF GAS LEAKS BY UNSUPERVISED NEURAL
NETWORK IN DISRIBUTION PIPELINES**

Campinas – SP
Janeiro, 2017

LUIZA BAPTISTA FERNANDES

DETECÇÃO E DIAGNÓSTICO DE VAZAMENTOS DE GÁS POR REDE NEURAL NÃO
SUPERVISIONADA EM TUBULAÇÃO DE DISTRIBUIÇÃO

DETECTION AND DIAGNOSIS OF GAS LEAKS BY UNSUPERVISED NEURAL
NETWORK IN DISTRIBUTION PIPELINES

Dissertação apresentada à Faculdade de Engenharia Química da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Engenharia Química.

Dissertation presented to the Chemical Engineering Faculty of the University of Campinas in partial fulfillment of the requirements for the degree of Master in the area of Chemical Engineering.

Orientadora: Profa. Dra. Ana Maria Frattini Fileti

ESTE EXEMPLAR CORRESPONDE À
VERSÃO DA DISSERTAÇÃO
DEFENDIDA PELA ALUNA LUIZA
BAPTISTA FERNANDES, E
ORIENTADA PELA PROFA. DRA. ANA
MARIA FRATTINI FILETI

Campinas – SP

Janeiro, 2017

Agência(s) de fomento e nº(s) de processo(s): CAPES, 33003017034P8

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

F391a Fernandes, Luiza Baptista, 1991-
Detecção e diagnóstico de vazamentos de gás por rede neural não supervisionada em tubulação de distribuição / Luiza Baptista Fernandes. – Campinas, SP : [s.n.], 2017.

Orientador: Ana Maria Frattini Fileti.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Química.

1. Redes neurais (Computação). 2. Gás - Vazamento. 3. FFT. 4. PCA. I. Fileti, Ana Maria Frattini, 1965-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Química. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Detection and diagnosis of gas leaks by unsupervised neural network in distribution pipelines

Palavras-chave em inglês:

Neural Networks (Computing)

Gas - Leakage

FFT

PCA

Área de concentração: Sistemas de Processos Químicos e Informática

Titulação: Mestra em Engenharia Química

Banca examinadora:

Ana Maria Frattini Fileti [Orientador]

Alberto Luiz Serpa

Luz Adriana Alvarez Toro

Data de defesa: 24-01-2017

Programa de Pós-Graduação: Engenharia Química

FOLHA DE APROVAÇÃO

Dissertação de Mestrado defendida por Luiza Baptista Fernandes e aprovada em 24 de janeiro de 2017 pela banca examinadora constituída pelos doutores:

Profa. Dra. Ana Maria Frattini Fileti

FEQ – UNICAMP

Prof. Dr. Alberto Luiz Serpa

FEM- UNICAMP

Profa. Dra. Luz Adriana Alvarez Toro

FEQ – UNICAMP

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica da aluna.

Dedico este trabalho a todos que direta ou indiretamente influenciaram sua composição, sejam eles professores, amigos, pai, noivo, irmão e, em especial, como sempre, a minha mãe.

AGRADECIMENTOS

Agradeço a minha família pelo apoio que sempre deu aos meus estudos. Especificamente, agradeço ao meu irmão Pedro por todo material que enviou sobre redes neurais.

Agradeço a minha mãe.

Agradeço ao meu noivo, por toda compreensão e por toda a ajuda nas pesquisas.

A minha orientadora professora Dra. Ana Maria Frattini Fileti, pela disposição em me orientar sempre com boas ideias e pelo cuidado demonstrado à pesquisa. Ao professor Dr. Flávio Vasconcelos da Silva pelas inúmeras aulas que me deixou assistir. A todos os colegas do Laboratório de Controle e Automação de Processos; ao Eng. Marcos Estevom pela ajuda em qualquer alteração que se fez necessária no sistema experimental utilizado.

A todos os professores que convivi e tive aula nos últimos dois anos na FEQ/Unicamp, sem exceção, todos de muita competência e conhecimento. Às amigas, que sempre me motivam.

Agradeço ao meu ex-orientador de iniciação científica, o Prof. Dr. José Carlos Cardoso Filho, responsável pelo laboratório de corrosão da FEQ/UFPA, por todo apoio e incentivo dado durante a minha graduação, graduação sanduíche e ingresso na pós-graduação.

A CAPES pelo apoio financeiro.

RESUMO

Vazamentos são em sua maioria imprevisíveis e indesejáveis. Dependendo da quantidade e do material em vazamento, representam um grande perigo ao ecossistema e efeitos financeiros negativos à instituição responsável. Assim, da importância da identificação da ocorrência de vazamentos, estudou-se um método para detecção e caracterização de vazamentos, de ar comprimido, em tubulação a partir do som produzido. Este método tem como suposição inicial que as respostas de vazamentos diferentes, ou de ausência de vazamentos, possuem padrões acústicos próprios, podendo, assim serem distinguidos um dos outros. Para a análise dos padrões, e identificação, a rede neural não supervisionada de Kohonen foi utilizada. Porém, primeiramente, a aquisição de dados foi realizada por microfones, acoplados à tubulação, obtendo dados no domínio do tempo. Cada simulação compreendeu dados em regime transiente e permanente. O domínio da frequência das amostragens contínuas foi calculado por Transformada Rápida de Fourier (FFT). O resultado de todos os dados experimentais, no domínio da frequência, foi uma matriz de dimensões de 5.475 x 6.599 (observações x dimensões). As 6.599 dimensões foram comprimidas a 200, pelo método de Análise de Componentes Principais (PCA), com 98% de representatividade. Os dados utilizados para treinamento das redes estavam, então, compreendidos na matriz resultante da compressão com dimensões de 5.475 x 200; testes usando dados não comprimidos para o treinamento da rede foram também realizados neste trabalho. A rede de Kohonen se mostrou uma forte ferramenta para a análise dos padrões acústicos tendo a vantagem de não necessitar de dados de saída do sistema para o seu treinamento.

PALAVRAS CHAVES:

Rede Neural de Kohonen, identificação de padrões, FFT, PCA, vazamento

ABSTRACT

Leaks are mostly unpredictable and undesirable. Depending on the quantity and on the material leaked, these situations represent a potential danger to the ecosystem and negative financial effects to the responsible institution. Thus, considering the importance of identifying the occurrence of leaks, a method was studied for detection and characterization of leaks of gas in pipes according to the sound produced. This method has as an initial assumption that the responses of each different leak, or absence of leak, have its own acoustic patterns, and so can be distinguished from one another. For analysing the features, Kohonen's unsupervised neural network was used. However, firstly, data acquisition was performed by microphones coupled to the pipeline to get the time domain acoustic response. Each experiment consisted of acquisitions of both transient and steady state responses. The frequency domain was calculated by Fast Fourier Transform (FFT). For results, the matrix size comprising data in terms of frequency was of 5.475 x 6.599 (observations x dimensions). The 6.599 dimensions were compressed to 200 using the Principal Components Analysis (PCA) with 98% of variance explained. Data used in the training of the Kohonen's net were comprised in the compressed matrix of dimensions of 5.475 x 200; tests using not compressed data to train the artificial neural network are also in this work. The net was found to be a strong tool to analyse and identify the acoustic patterns. Besides, it was advantageous not to necessarily attribute output to the training data of the network.

Key-words:

Kohonen Self-Organizing Map, features identification, FFT, PCA, leakage.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 2.1 Ilustração de uma análise de Componentes Principais, todas as dimensões são mantidas [20]. À direita, nas dimensões originais; à esquerda, em termos das componentes principais | 30 |
| Figura 2.2 Esquema da organização de vizinhança dos neurônios na camada de saída. (a) hexagonal, (b) retangular. Adaptado de [27]..... | 35 |
| Figura 2.3 Topologias diferentes para o mapa auto-organizável de Kohonen: (a) formato plano; (b), cilíndrico e (c) toróide. Adaptado de [27]..... | 36 |
| Figura 2.4 Inicializações dos vetores pesos, em azul. À esquerda, randômica; à direita, linear. Dados de treino randômicos simbolizados por “+”. Gerados no <i>software Matlab</i> | 37 |
| Figura 2.5 Mapa de Rede Auto-Organizável de Kohonen. Adaptado de [25] | 38 |
| Figura 2.6 Ordenação dos pesos com o aumento de iterações de treinamento. Adaptado de [26] | 42 |
| Figura 2.7 Esquema de geração da matriz U em vizinhança hexagonal de neurônios 3x3. Adaptado de [31] | 43 |
| Figura 3.1 Sistema utilizado para simulações de vazamentos (LCAP/FEQ/Unicamp) | 47 |
| Figura 3.2 Resposta de Frequência do microfone modelo XCM-9767 [5] | 48 |
| Figura 3.3 Ambiente virtual do Labview para aquisição de dados, página “Front Panel” | 50 |
| Figura 3.4 Esquema da aquisição e tratamento de dados. Ramo principal: metodologia investigada; ramo secundário, testes de metodologias alternativas..... | 58 |
| Figura 4.1 (a) Amplitude capturada ao longo do tempo em ocorrência de vazamento em F1; (b) Variação da amplitude nos microfones com a ocorrência do vazamento em F1; (c) Amplitude capturada ao longo do tempo em ocorrência de vazamento em F2; (d) Variação da amplitude nos microfones com a ocorrência do vazamento em F2 | 62 |
| Figura 4.2 Vazamento em F1 em tempos aproximados diferentes. (a) antes de simular vazamento; (b) em aproximadamente 20 s; (c) em aproximadamente em 30 s; (d) em aproximadamente em 60 s | 64 |
| Figura 4.3 Teste Cumulativo de Variância dos dados de frequência em termos de componentes principais | 66 |
| Figura 4.4 (a) Todos os dados experimentais em 3 dimensões; (b) Região ampliada | 67 |
| Figura 4.5 Dados em 3 dimensões de dados de apenas um experimento: (a) De vazamento em F1; (b) de vazamento em F5 | 68 |

| | |
|---|----|
| Figura 4.6 Neurônios treinados e os dados de treinamento padronizados, em termos de suas componentes principais (C. P.)..... | 72 |
| Figura 4.7 Representação dos Neurônios em matriz U | 73 |
| Figura 4.8 Matriz U evidenciando ativação dos neurônios a microfones em posições distintas. (a) Matriz U evidenciando neurônios ativados por dados de treino; agrupados visualmente em (b). (c) Matriz U evidenciando neurônios ativados por dados de teste; agrupados visualmente em (d). Vermelho: m1; verde: m2; laranja: m6; azul: m11 e amarelo: m13..... | 74 |
| Figura 4.9 Matriz U evidenciado ativação dos neurônios aos dados de m2 apenas. (a) Dados de treino; (b). Dados de teste. Situações sem vazamentos- preto: sem anormalidades; rosa escuro: batidas no botijão; azul escuro: batidas na tubulação. Com vazamentos- vermelho: em F1 (1 mm); verde: em F2 (0,5 mm); azul: em F3 (2 mm); laranja: em F4 (1 mm); amarelo: em F5 (4 mm); ciano: em F6 (3 mm); e cinza: em F7 (0,5 mm) | 75 |
| Figura 4.10 Matriz U evidenciado ativação dos neurônios aos dados de m13 apenas. (a) Dados de treino; (b). Dados de teste. Situações sem vazamentos- preto: sem anormalidades; rosa escuro: batidas no botijão; azul escuro: batidas na tubulação. Com vazamentos- vermelho: em F1 (1 mm); verde: em F2 (0,5 mm); azul: em F3 (2 mm); laranja: em F4 (1 mm); amarelo: em F5 (4 mm); ciano: em F6 (3 mm); e cinza: em F7 (0,5 mm) | 76 |
| Figura 4.11 Neurônios treinados e os dados de treinamento de m1 padronizados, em termos de suas componentes principais (C. P.)..... | 77 |
| Figura 4.12 Representação dos neurônios em matriz U com dados apenas de m1 | 78 |
| Figura 4.13 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27) | 79 |
| Figura 4.14 Neurônios treinados e os dados de treinamento padronizados de m2, em termos de suas componentes principais (C. P.)..... | 81 |
| Figura 4.15 Representação dos neurônios em matriz U com dados apenas de m2..... | 82 |
| Figura 4.16 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais | |

| | |
|---|----|
| similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27) | 83 |
| Figura 4.17 Neurônios treinados e os dados de treinamento padronizados de m6, em termos de suas componentes principais (C. P.)..... | 85 |
| Figura 4.18 Representação dos neurônios em matriz U com dados apenas de m6 | 86 |
| Figura 4.19 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27) | 87 |
| Figura 4.20 Neurônios treinados e os dados de treinamento padronizados de m11, em termos de suas componentes principais (C. P.)..... | 90 |
| Figura 4.21 Representação dos neurônios em matriz U com dados apenas de m11 | 90 |
| Figura 4.22 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27) | 92 |
| Figura 4.23 Neurônios treinados e os dados de treinamento padronizados de m13, em termos de suas componentes principais (C. P.)..... | 95 |
| Figura 4.24 Representação dos neurônios em matriz U com dados apenas de m13 | 95 |
| Figura 4.25 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no | |

botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27)97

Figura 4.26 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com interpolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7 99

Figura 4.27 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com extrapolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7 101

Figura 4.28 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com extrapolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7 103

Figura 4.29 Matriz U com dados de m1 com amplitudes em frequências de 100 a 1.600 Hz105

Figura 4.30 Matriz U com dados de m13 com amplitudes em frequências de 100 a 1.600 Hz(a) Evidenciando neurônios ativados pelos dados de treino; (b) Evidenciando neurônios ativados pelos dados de teste 106

Figura 4.31 Matriz U com dados de m1 com amplitudes em tempo de simulação de aproximadamente 60 s. (a) Evidenciando neurônios ativados pelos dados de treino; (b) Evidenciando neurônios ativados pelos dados de teste 108

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 3.1 Distâncias dos furos com relação a entrada do gás comprimido na tubulação..... | 47 |
| Tabela 3.2 Distância dos microfones com relação a entrada do gás comprimido na tubulação | 49 |
| Tabela 3.3 Experimentos Conduzidos | 59 |
| Tabela 4.1 Redes realizadas..... | 70 |
| Tabela 4.2 Parâmetros de Treinamento das Redes | 71 |
| Tabela 4.3 Parâmetros de treinamento das redes de interpolação e de extrapolação | 102 |

LISTA DE ABREVIATURAS E SIGLAS

C.P. 1- componente principal com maior autovalor, primeira componente principal

C.P. k - componente principal com menor autovalor

C.P. 2 - segunda componente principal

DFT- Transformada Discreta de Fourier

FFT- Transformada Rápida de Fourier

PCA- Análise de Componentes Principais

PCV - percentual cumulativo de variância

SOM - redes não supervisionadas, *self-organizing maps*

LISTA DE SÍMBOLOS

Letras latinas

c - neurônio vencedor

d_{ci} - distância Euclidiana do neurônio vencedor c ao neurônio i

f - frequência

F_n - furos na tubulação

h_{ic} - função de vizinhança em relação ao neurônios vencedor c e neurônios vizinhos i

$h_{ic\ bubble}$ - função “bubble” de vizinhança

$h_{ic\ gaussian}$ - função “gaussian” de vizinhança, gaussiana

$h_{ic\ cutgauss}$ - função “cutgauss” de vizinhança

i - neurônio i da vizinhança do neurônio vencedor

$j(x)$ - função discriminante

k - número original de dimensões

m - valor da frequência em que o sinal é investigado

mn - microfones na tubulação

n - valor discreto do tempo

N - Número de amostras de entrada no domínio do tempo e número de frequências avaliadas de saída, domínio da frequência

t - tempo, ou número de iterações no treinamento da rede neural

u - número de componentes principais utilizadas

$X(f)$ - sinal no domínio da frequência f

$X(t)$ - sinal no domínio do tempo t

w_i - vetores pesos, neurônios

$w_i(t + 1)$ - vetores pesos atualizados

$w_i(t)$ - vetores pesos anteriores à atualização

x - entradas de um processo

$x(n)$ - sinal discreto no domínio do tempo

$X(m)$ - sinal discreto no domínio da frequência

y - saídas de um processo

Grego

$\alpha(t)$ - função taxa de aprendizado

$\alpha_0(t)$ - função taxa de aprendizado inicial

$\alpha(t)_{inversa}$ - função “inversa” de taxa de aprendizado

$\alpha(t)_{power}$ - função “power” de taxa de aprendizado

$\alpha(t)_{linear}$ - função “linear” de taxa de aprendizado

γ_z - autovalor z correspondente à componente principal z usada

σ_t - raio da vizinhança

SUMÁRIO

| | |
|--|----|
| 1 INTRODUÇÃO..... | 20 |
| 1.1 Objetivos do Projeto | 23 |
| 2 FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA..... | 25 |
| 2.1 Método Acústico de Detecção de Vazamento..... | 25 |
| 2.2 Processamento de Sinais..... | 26 |
| 2.3 Transformada Rápida de Fourier (FFT) | 27 |
| 2.4 Análise de Componentes Principais (PCA)..... | 29 |
| 2.4.1 Interpretação Geométrica de PCA..... | 30 |
| 2.4.2 A Busca pelas Componentes Principais | 31 |
| 2.5 Redes Neurais | 32 |
| 2.5.1 Redes Neurais Supervisionadas..... | 33 |
| 2.5.2 Redes Neurais Não Supervisionadas | 33 |
| 2.5.2.1 Redes de Kohonen | 34 |
| 1) Inicializações da Rede | 36 |
| 2) Treinamentos da Rede | 37 |
| 3) Mapeamento dos Neurônios: Erros de Quantização e Topográfico..... | 41 |
| 2.5.2.2 Matrix U | 42 |
| 2.5.3 Conclusões sobre Redes Neurais para a Proposta de Pesquisa | 44 |
| 3 METODOLOGIA..... | 46 |
| 3.1 Sistema montado de Simulação dos Vazamentos..... | 46 |
| 3.2 Dispositivos para aquisição | 47 |
| 3.2.1 Microfones..... | 48 |
| 3.2.2 Pré-amplificador | 49 |
| 3.2.3 Conversor Analógico/Digital e CPU | 49 |
| 3.3 Aquisição de Dados no Domínio do Tempo e da Frequência | 50 |
| 3.3.1 Aquisição dos dados no domínio do tempo..... | 51 |

| | |
|--|----|
| 1) Definição de Valores de Parâmetros Para a Aquisição | 51 |
| 2) Filtro..... | 51 |
| 3) Armazenamento de Dados | 52 |
| 3.3.2 Cálculo das Amplitudes no Domínio da Frequência | 52 |
| 3.4 Redução da Dimensionalidade e Redes Neurais | 52 |
| 3.4.1 Investigações com Dados no Tempo | 53 |
| 3.4.2 Investigações com Dados no Domínio da Frequência..... | 53 |
| 3.4.3 Redução da Dimensionalidade dos Dados em Frequência..... | 53 |
| 3.4.4 Criação de Redes para a Identificação, Caracterização e Localização de Vazamentos..... | 54 |
| 1) Redes Treinadas com Dados em Dimensão Reduzida por PCA..... | 54 |
| 1.1) Criação da Rede Neural..... | 54 |
| 1.2) Teste de Interpolação | 55 |
| 1.3) Testes de Extrapolações | 56 |
| 2) Redes Treinadas com Dados diretamente no Domínio da Frequência..... | 56 |
| 2.1) Redução da Dimensionalidade por Exclusão de Frequências na criação da Rede..... | 57 |
| 2.2) Redução da Dimensionalidade escolhendo apenas Dados em Regime Permanente..... | 58 |
| 3.5 Planejamento dos Experimentos..... | 59 |
| 4 RESULTADOS E DISCUSSÕES..... | 60 |
| 4.1 Resultados de Aquisição..... | 60 |
| 4.2 Domínio do tempo | 61 |
| 4.3 Domínio da Frequência | 63 |
| 4.4 Análise de Componentes Principais | 65 |
| 4.5 Rede Neural de Kohonen..... | 68 |
| 4.5.1 Redes Treinadas com Dados Comprimidos por PCA | 69 |
| 1) Construção da Rede 1: Todos os Dados..... | 71 |

| | |
|---|-----|
| 2) Construção da Rede 2: Dados do Microfone 1 | 77 |
| 3) Construção da Rede 3: Dados do Microfone 2 | 81 |
| 4) Construção da Rede 4: Dados do Microfone 6 | 85 |
| 5) Construção da Rede 5: Dados do Microfone 11 | 89 |
| 6) Construção da Rede 6: Dados do Microfone 13 | 94 |
| 7) Conclusão sobre as Redes 2 a 6 | 98 |
| 8) Respostas de Interpolação e de Extrapolação na Rede de Kohonen | 98 |
| 4.5.2 Testes de Redes Não Supervisionadas Sem PCA..... | 104 |
| 1) Com Amplitudes em Frequências Aleatórias..... | 104 |
| 2) Com Amplitudes em Regime Permanente | 107 |
| 4.6 Propostas de Uso em Montagem Física em Sistema Equivalente | 109 |
| 4.7 Sugestões para Trabalhos Futuros | 109 |
| 5 CONCLUSÕES | 111 |
| REFERÊNCIAS | 116 |
| APÊNDICE A | 119 |
| APÊNDICE B..... | 125 |
| APÊNDICE C..... | 133 |
| APÊNDICE D | 135 |
| APÊNDICE E..... | 137 |

1 INTRODUÇÃO

Muitos dos vazamentos são caracterizados por perda indesejável e imprevisível de materiais e, dependendo das características do material perdido e da extensão do vazamento, são potencialmente perigosos. Assim, ao lidar com a distribuição e com o armazenamento de materiais através de, por exemplo, tubulações e de tanques de armazenamento, deve-se prestar atenção aos fatores de segurança.

Um exemplo próximo ao cotidiano é o gás de cozinha, frequentemente presente na rotina urbana, este é um gás inflamável e intoxicante. Seu vazamento é preocupante e um exemplo de medidas de identificações de seu vazamento é realizado na própria refinaria: a de se acrescentar odores artificiais ao gás. Assim, sua fuga eventual é mais rapidamente identificada.

Motivos de vazamento são geralmente associados a falhas na estrutura da tubulação ou do recipiente de armazenamento. Tais falhas podem ter diversas origens como fadiga por pressão, corrosão, defeito de fábrica etc e são muitas vezes imprevisíveis e os danos podem tomar grandes proporções negativas financeiras, ambientais e sobre a saúde.

Diante da questão, diversos trabalhos têm como foco de estudo seja a prevenção de vazamentos, ou a sua identificação. Em prevenção, estudam-se a estrutura dos materiais, a localização de pontos mais suscetíveis a falhas estruturais, métodos de se evitar corrosões e fadigas etc. Enquanto que, em identificação, estudos focam em como se detectar ou caracterizar um vazamento uma vez que haja suspeita. Ou seja, de modo geral, buscam-se formas confiáveis de se evitar e de se detectar vazamentos, posto que estes representam materiais perdidos, desviados do seu destino final, indesejados e inesperadamente espalhados sobre uma área.

O método de detecção de vazamento estudado neste trabalho é o método acústico, este investiga o comportamento das ondas mecânicas acústicas geradas pelo sistema com a ocorrência, ou não, de algum defeito na tubulação que permita fuga de material interno.

Assim, no estudo de detecção e caracterização de vazamentos, o método acústico possui a premissa básica de que o vazamento, ao ocorrer, produz um som próprio de acordo com suas características. Este som, capturado por microfones, pode ser analisado no domínio do tempo como a somatória resultante de todas as ondas acústicas geradas pelo sistema em cada instante.

Além do método acústico, existem métodos baseados em balanços de massa do sistema, ou de volume, no entanto, dificilmente indicam a localização do vazamento, assim como métodos baseados na pressão [1]. Porém, variações de temperatura e de pressão ao longo da tubulação dificultam os balanços, pois provocam alterações das características dos fluidos necessárias para os cálculos (como variações na massa específica, por exemplo).

Além disso, técnicas de detecção por modelagem de um sistema com vazamento são encontradas na literatura desde décadas atrás. A exemplo, em 1988, foi proposto modelo conforme balanços de energia e de massa para fluxos unidimensionais hidrodinâmicos transientes com a limitação de ser para fluidos incompressíveis, e trabalhos anteriores a esse possuíam a limitação de serem para fluxo laminar [2]. Num outro estudo, de 2012, com vazamento de óleo, fez-se um estudo do sistema a partir de circuito equivalente ao tubo com vazamento para gerar um modelo matemático numa aproximação válida [3]. Desta forma, há muitas alternativas já exploradas para análise de vazamentos.

O ideal, para a detecção de vazamentos com o menor número possível de alarmes falsos, é a utilização em conjunto das técnicas existentes. Este trabalho se foca apenas no monitoramento acústico no sentido de estudá-lo e de aprimorá-lo.

Para uma análise mais detalhada do som capturado pelo microfone, o sinal é decomposto para o seu domínio de frequência, no qual as características do sinal podem ser também analisadas em termos de frequência. A técnica utilizada para o cálculo das componentes em termos de frequência foi a Transformada Rápida de Fourier (FFT).

Assim, dados foram capturados por microfones ao longo do tempo: desde o início do vazamento simulado até um tempo final de simulação. Pressupõem-se que o tempo de aquisição de dados seja longo o suficiente para que o comportamento acústico já tenha atingido um regime permanente. Assim, os dados representam tanto o regime transiente quanto dados em regime permanente.

Da aquisição de dados, no domínio do tempo houve observações de amplitudes ao longo do tempo pelo tempo determinado na metodologia; E, no domínio da frequência, houve observações de amplitudes nas frequências investigadas. Porém, por se tratar de uma amostragem continuada, dados em frequência variam de um instante a outro. Dessa forma, cada amplitude calculada para uma dada frequência, dentro do intervalo investigado de frequência, representa uma dimensão, ou seja, uma variável com observações experimentais que variam ao longo do tempo de amostragem.

Assim, observações em muitas dimensões são geradas ao se ter a decomposição do som do domínio do tempo para o de frequência. Desta forma, uma alternativa para comprimir as observações do domínio da frequência, de modo a diminuir as dimensões sem muita perda de informação, é a utilização do método estatístico de Análise de Componentes Principais (PCA).

Este método, PCA, de um modo geral, busca as componentes principais dos dados a partir de suas dimensões originais. Tais componentes principais são os vetores, ortogonais um ao outro, em que em suas direções se obtém o máximo de variância dos dados originais. Assim, as componentes principais são geradas com a ordem: a primeira componente principal é a direção nos dados que possui maior variância, a segunda componente é a segunda direção, ortogonal à primeira, com maior variância e assim por diante até se obter um número de componentes principais iguais ao número de dimensões nos dados de origem.

Assim, o número máximo de componentes principais é igual ao número de dimensões originais. Neste método, porém, muitas vezes, uma representação significativa dos dados originais é alcançada com uma dimensionalidade menor do que a original, pois a variância contida em cada direção diminui conforme o número de componentes aumenta. As direções com pouca variância não adicionam muita informação e podem ser ignoradas sem grandes perdas de representação. O PCA não diminui o número de observações, apenas a dimensão dos dados observados.

Após a compressão dos dados, propõe-se que o resultado do PCA seja entrada para outra técnica: a rede neural.

As redes neurais podem ser supervisionadas ou não supervisionadas. As supervisionadas possuem a desvantagem de necessariamente ter que se atribuírem saídas numéricas aos dados de entrada de treinamento da rede neural. No caso de vazamentos, esta atribuição se torna complexa e pode não convir, tornando-se mais difícil ainda com o aumento da complexidade do vazamento. Assim, este trabalho utiliza de uma rede neural não supervisionada, a rede de Kohonen. Esta rede estuda os padrões dos dados de entrada, se vazamentos diferentes possuem padrões diferentes, a rede consegue distingui-los mesmo que não seja fornecido nenhum dado de saída.

A saída da rede é, mais especificamente, o mapeamento de neurônios de forma que estes representem padrões dos dados de entrada. Assim, uma forma de identificar o sistema é que, por exemplo, em situações de vazamentos com certas características (tamanho de diâmetro ou localização, por exemplo), apenas os neurônios que mapeiam este padrão

específico sejam ativados. Desta forma, numa situação prática de monitoramento de tubulações de gás por sensores acústicos, se os neurônios conhecidamente associados a um padrão de vazamento forem ativados pelos dados provindos de um sistema, um alarme poderá avisar o evento.

Para a visualização do resultado do treinamento não supervisionadas, é proposto a utilização de matriz U , ou *u-matrix*, do inglês. Nessa matriz, independentemente da dimensionalidade dos dados, as investigações podem ser visualizadas em 2-D.

Desta forma, o presente trabalho foca na identificação e na caracterização, dimensão do vazamento e localização, de vazamentos ao longo de tubulações com a sua ocorrência. O diagnóstico é baseado no som característico do sistema em que o sinal será decomposto para o domínio da frequência e tem sua dimensão reduzida para que a rede neural possa fornecer meios de identificação e localização dos vazamentos investigados através dos padrões gerados. Então, procura-se investigar se as técnicas propostas são úteis para diagnóstico caso haja suspeita de vazamento em tubulações de distribuição.

1.1 Objetivos do Projeto

Diante do desafio de identificar o vazamento assim que haja sua ocorrência, o presente trabalho traçou objetivos a serem efetivados, analisados e discutidos conforme progresso do estudo.

1) Objetivo Geral

Comprovar que os métodos matemáticos propostos, utilizados computacionalmente em sequência, podem ser utilizados para monitoramento acústico eficiente capaz de acusar a ocorrência de eventos anormais em tubulações de distribuição e caracterizar esta ocorrência.

2) Objetivos Específicos

Obtenção de dados no sistema experimental para simulação de vazamentos montado no Laboratório de Controle e Automação de Processos da FEQ/UNICAMP;

Criar rotina em *LabView* para aquisição de dados no domínio do tempo, e então, no domínio da frequência por Transformada Rápida de Fourier (FFT); determinar parâmetros para aquisição eficiente dos dados;

Criar rotinas no *Matlab* para a análise dos dados em componentes principais, PCA;

Criar rotinas no *Matlab* de treinamentos da rede neural proposta;

Avaliar qualitativamente a eficiência da rede para identificar vazamentos;

Avaliar a eficiência da identificação da dimensão de vazamentos e da sua localização com a rede não supervisionada de Kohonen, tendo o método acústico para a aquisição e os dados observados em frequência comprimidos pelo método PCA como entrada de treinamento da rede;

Avaliar métodos alternativos para diminuição de dados (sem PCA);

Avaliar respostas de extrapolações e de interpolações da rede não supervisionada;

Propor montagem eficiente de microfones, em termos de espaçamentos, em tubulação equivalente de acordo com os resultados obtidos.

2 FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA

Nesta seção, trabalhos encontrados na literatura e uma breve fundamentação teórica dos métodos propostos para a identificação e diagnóstico de ocorrências de vazamento serão expostos.

2.1 Método Acústico de Detecção de Vazamento

O método acústico utilizado na detecção de vazamento é caracterizado pela utilização de sensores acústicos, como microfones, na investigação do sistema. Estes são capazes de perceber diferenças nas ondas acústicas quando um vazamento começa a ocorrer.

Ao ocorrer um vazamento, o sistema que comporta o gás é modificado. Junto com outras propriedades alteradas pelo vazamento, que também podem ser utilizadas para diagnosticar um vazamento (como a pressão do sistema, ou a quantidade de massa etc), há o som gerado pelo sistema. Tal alteração do som, mesmo que muito baixa e despercebida à sensibilidade sonora humana, pode ser capturada pelo microfone, investigada e usada como um parâmetro de estudo para identificação do vazamento.

Técnicas para medição acústica externa, e assim evitar alterações na estrutura da tubulação ao acoplar sensores ou evitar que os sensores percam eficiência com o contato direto com o sistema, são encontradas na literatura. Por exemplo, a utilização de ultrassom ou de termografia. De fato, equipamentos de ultrassom são muito utilizados principalmente na medicina, porém também têm utilidade em indústrias químicas [4]. No entanto, técnicas de ultrassonografia são mais comumente utilizadas com líquidos em escoamento.

O método acústico para detecção de vazamento já se mostrou bastante eficiente em trabalhos como [5, 6, 7] utilizando microfones acoplados ao sistema com ar comprimido em vazamento. Nestes trabalhos também se mostrou claro que tamanhos de orifícios diferentes influenciavam na resposta da onda sonora. Esta característica pode ser usada para estimar o tamanho do orifício.

Em [5], microfones com mais de 30 m de distância da fonte de um vazamento de 0,5 mm de diâmetro detectaram perturbações geradas pelo escapamento de gás comprimido em pressão inicial de 1 kgf/cm².

Ao se ter o comportamento acústico de um sistema, pode-se utilizar de conhecimentos físicos da onda acústica para estimar a localização do vazamento. Em [5], é proposto um cálculo simples em que a localização do furo causador do vazamento pode ser calculada, tendo-se conhecimento da velocidade da onda acústica no meio, a partir do tempo que o som demora para chegar em dois sensores acoplados à tubulação em diferentes posições e da distância entre eles.

Nos trabalhos citados [1, 5, 6, 7], o método acústico se prova capaz de disponibilizar dados contendo sensibilidade suficiente para diagnosticar alterações no sistema mesmo distante do sensor, além de providenciar dados em que a localização pode ser estimada assim como a dimensão do rompimento da tubulação.

Os sinais explorados no domínio do tempo fornecem o comportamento do sistema antes, no momento e após que as perturbações ocorrem. Sua análise direta pode fornecer informações sobre anormalidades ocorrendo no sistema como alterações na amplitude média da onda, ou modificações no seu comportamento.

2.2 Processamento de Sinais

Processamentos de sinais podem ser divididos em analógicos ou digitais. Processamentos de sinais analógicos envolvem uma entrada de sinal contínuo no tempo. Sinais analógicos são processados por dispositivos como amplificadores, transistores-dispositivos ativos- ou, por dispositivos como resistores, capacitores, indutores- dispositivos passivos. A saída do sistema também é contínua no tempo.

Processamento de sinais digitais, ou sinais discretos, possuem como entrada uma sequência de números binários que é processada, por exemplo, por computadores, dispositivos eletrônicos e outros, os quais possuem uma saída também binária [8].

Sinais analógicos podem ser transformados em digitais (e vice-versa) para que possam ser transmitidos e processados digitalmente. A conversão de um sinal analógico para digital possui as três seguintes etapas: a amostragem, a quantização e a codificação, respectivamente. A amostragem corresponde à medição da amplitude do sinal analógico em espaços de tempo iguais e definidos. A frequência de amostragem deve ser ao menos duas vezes maior do que a frequência em análise para que não ocorra o efeito *aliasing*. Neste fenômeno, é possível que ondas de frequências diferentes, amostradas em uma frequência mal ajustada, menor do que o dobro da maior frequência, se tornem indistinguíveis [9, 10].

A etapa de quantização consiste em associar os valores amostrados aos seus respectivos níveis de quantização, estes são conjuntos finitos de valores que o sinal digitalizado poderá assumir. A quantidade de níveis depende do número de bits do conversor analógico digital. Em n bits, haverá 2^n níveis quantização. O valor amostrado pertencerá ao nível mais próximo de seu valor para que o erro de quantização final seja o menor possível [1, 10]. Quanto maior o número de bits, mais exata é a conversão por possuir mais níveis de quantização.

Por fim, a codificação consiste em associar os níveis quantizados a números binários. Uma forma padrão deste processo é o conhecido como Modulação por Códigos de Pulso [1]. O número binário final é o sinal completamente digitalizado.

2.3 Transformada Rápida de Fourier (FFT)

Ao se amostrar sinais no tempo, a onda capturada por sensores é o somatório resultante de todas as ondas de frequências e amplitudes diferentes que houver no sistema em cada instante de tempo. Desta forma, decompor a onda no domínio da frequência faz com que a análise do sinal se torne mais específica. A exemplo, na dissertação [1], componentes sonoras de frequências diferentes surgem no sistema conforme a situação experimental, enquanto que, no domínio do tempo, dificilmente se perceberia a mudança. Um método de se transformar um sinal do domínio do tempo para o domínio da frequência é utilizando a Transformada Rápida de Fourier.

A Transformada Rápida de Fourier é o resultado de implementações de algoritmos desenvolvidos para se resolver mais rapidamente a Transformada Discreta de Fourier (DFT). Em 1965, um artigo foi publicado por Cooley e Tukey introduzindo a FFT [11, 12].

A Transformada de Fourier é uma ferramenta matemática que tem como entrada o sinal determinado no domínio do tempo e, como saída, o mesmo sinal decomposto no domínio da frequência. Ou seja, torna aparente o comportamento do sinal em termos de frequência. Esta transformada representa sinais como uma somatória de senos e cossenos [13].

A origem da Transformada Discreta de Fourier está na transformada contínua de Fourier $X(f)$,

$$\int_{-\infty}^{+\infty} X(t)e^{-j2\pi ft} dt \quad (2.1)$$

onde $X(f)$ é a decomposição do sinal $X(t)$, no domínio do tempo (t), para o domínio da frequência (f).

No caso de se trabalhar com dados digitalizados, discretos, a transformada de Fourier pode ser definida conforme $X(m)$ a seguir.

$$X(m) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi mn/N} \quad (2.2)$$

Ou, pela relação de Euler:

$$X(m) = \sum_{n=0}^{N-1} x(n)[\cos(j2\pi mn/N) - j\sin(j2\pi mn/N)] \quad (2.3)$$

onde,

$X(m)$: é o valor da decomposição do sinal discreto; se não houver componentes na respectiva frequência, o valor da somatória da eq. (2.3) será zero em real e em imaginário;

m : é o respectivo valor de frequência em que o sinal é investigado; $m=0,1,2,3,\dots,N-1$;

$x(n)$: o sinal de entrada no domínio do tempo;

n : valores do tempo, $n=0,1,2,3,\dots,N-1$;

N : o número de amostras de entrada no domínio do tempo e o número de frequências avaliadas para a saída, domínio da frequência.

Para se realizar a DFT é necessário realizar N^2 cálculos envolvendo números complexos, pois é o somatório de n variando de 0 à $N-1$ na eq.(3) para cada valor de m que também varia de 0 à $N-1$.

Na Transformada Rápida de Fourier, a quantidade de cálculos se reduz de modo a exigir um esforço muito menor para a resolução da transformada [12].

Especificamente para a análise de vazamentos, em [14], por exemplo, a análise no domínio da frequência apresentava novas componentes quando havia ocorrência de vazamento. Assim, a FFT é uma técnica geradora de dados que são diferentes conforme as situações com vazamento, ou sem vazamento, ocorrem.

2.4 Análise de Componentes Principais (PCA)

Em dados experimentais, quando muitas variáveis estão em análise, é possível que algumas sejam correlacionadas, ou que não adicionem muita informação à investigação. A análise de componentes principais de um conjunto de dados possui como expectativa diminuir o número de variáveis, em um conjunto menor e não correlacionado.

De maneira geral, o método de Análise de Componentes Principais (do inglês, *Principal Component Analysis*, PCA) atua fazendo combinações lineares de tal forma que a variância das observações se concentre na sua maioria em um número menor de dimensões, ortogonais entre si, de modo que, ao se ignorar as dimensões em cuja direção os dados não variem muito, não se perca muita informação. Assim, procede-se a diminuição de dimensões.

O método PCA foi introduzido num artigo de 1901 de Karl Pearson [15], mas a técnica estatística é continuamente desenvolvida desde então [16]. Algumas das utilidades práticas do PCA é a redução da dimensionalidade dos dados, número de variáveis, em monitoramento de sistema industriais para detecção de falhas [17]; reconhecimento de padrões em faces e compressão de imagens [18]; análises de sinal. Há na literatura também trabalhos que fizeram uso da técnica em análises biológicas como em análise de fadiga muscular [19], sendo inclusive usada em conjunto com FFT.

Na tese [5], PCA foi utilizada para diminuir a dimensionalidade dos dados gerados no domínio da frequência, calculados por FFT, de sinais capturados por microfones em um sistema em que vazamentos eram simulados. A análise de PCA obteve sucesso em diminuir a dimensão de dados de uma matriz de dados com dimensões 46.970x2.305 (46.970 sinais de amplitudes em frequências diferentes com 2.305 observações) com representatividade praticamente íntegra em uma matriz em termos de componentes principais de dimensões 2306x2305 (2.306 componentes principais com 2.305 observações). Para os propósitos da tese supracitada, mantiveram-se apenas 200 componentes principais.

Em um conjunto de dados com k dimensões, k componentes principais podem ser calculadas. As componentes principais são matematicamente as direções, arranjadas ortogonalmente, em que os dados apresentam maior variância. Elas são organizadas em ordem decrescente. Assim, a componente principal 1 (C.P. 1) é o eixo, ou vetor, cuja a direção é aquela em que os dados apresentam variância máxima e a componente principal k (C.P. k) é a direção de variância mínima.

Em muitos casos, as componentes principais iniciais já concentram boa parte da variância total. Assim, ao se manter apenas as primeiras componentes, pouca informação se perde, por compreender pouca variância total nas componentes deixadas de lado. Então, o conjunto de dados passa a ser representado em menos dimensões, em termos das componentes principais suficientes para representar a variância do sistema. A diminuição de dimensões facilita, entre outros âmbitos, a análise de padrões nos dados.

2.4.1 Interpretação Geométrica de PCA

Com base na Figura 2.1 à esquerda, 50 observações foram feitas com relação a duas variáveis x e y [20]. O gráfico do mesmo dado experimental em termos de suas componentes principais é a Figura 1 à direita. Como proposto pelo método, a componente principal 1 (C.P. 1) é a direção em que os dados apresentam mais variância e a segunda (C.P. 2), menor.

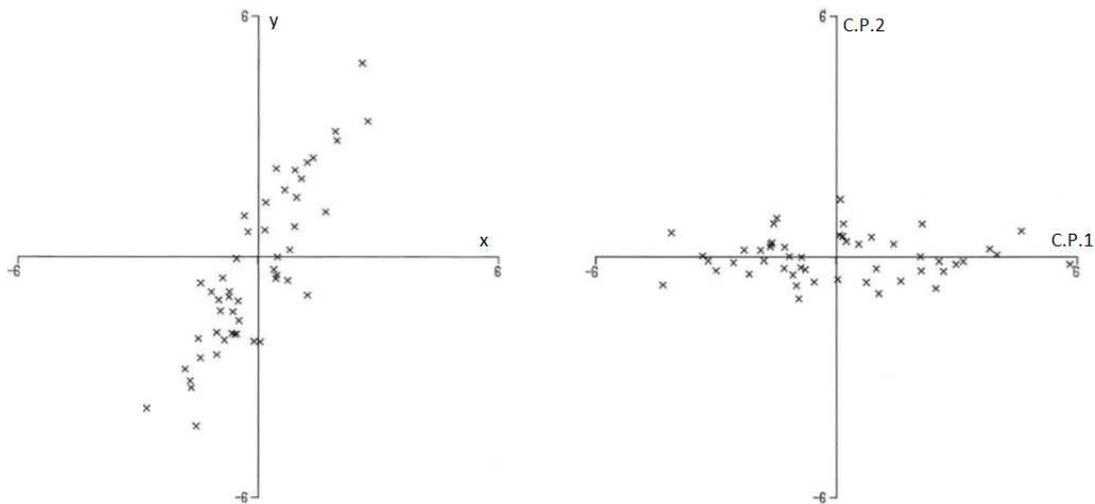


Figura 2.1 Ilustração de uma análise de Componentes Principais, todas as dimensões são mantidas [20]. À direita, nas dimensões originais; à esquerda, em termos das componentes principais

De acordo com a Figura 2.1, ao se preservar o número de dimensões original, duas dimensões, o método age apenas ao gerar rotação ou translação para novos eixos, as componentes principais. Se se optasse por reduzir dimensões no exemplo, apenas a componente principal de maior variância restaria, a C.P.1. A representação gráfica de apenas uma dimensão conteria todos os valores sobre o eixo C.P.1 considerando os valores observado

em C.P. 2 iguais a zero, inexistentes. Assim, C.P.2 seria a dimensão deixada de fora. Enquanto que, se se optasse por retirar uma das dimensões no gráfico original (eixo x ou eixo y da figura à direita da Figura 2.1), uma quantidade maior de variância se perderia e, possivelmente, prejudicaria, ou impossibilitaria a análise.

A escolha do número de dimensões que pode ser ignorada é de acordo com a variância que a direção da dimensão contém. Quanto menor a variância, menos informação a dimensão contém e mais facilmente pode ser desconsiderada.

2.4.2 A Busca pelas Componentes Principais

Antes da análise de componentes principais, os dados passam por um pré-tratamento. Cada variável é padronizada para possuir média de 0, ou seja, diminuir de cada valor a média observada de observações para a respectiva variável. Assim, a variação é avaliada em relação a média. Em seguida, as observações são divididas pelo seu desvio padrão para que fiquem também com desvio padrão padronizado em 1 para que variáveis com altas variância não prevaleçam sobre outras [17].

O mecanismo para se buscar as componentes principais consiste no cálculo dos autovetores, e seus respectivos autovalores, da matriz de covariância dos dados. De fato, as direções das componentes principais possuem as mesmas direções dos autovetores calculados. Os autovalores são os valores correspondentes às variâncias na direção do autovetor. As componentes principais são então organizadas conforme os valores acusados de autovalores para cada autovetor: a componente principal 1 possui a direção do autovetor da matriz de covariância que possui maior autovalor e assim por diante. Os cálculos específicos com exemplos do método podem ser encontrados em livros como [17, 20].

Uma vez encontrados os autovetores da matriz de covariância, os valores originais observados, respectivos às dimensões iniciais, passam pela transformação linear necessária para se obter os valores das observações correspondentes às componentes principais.

A escolha de quantas componentes principais são necessárias para se representar os dados, dentre outras análises, pode ser utilizado o teste Percentual Cumulativo da Variância (PCV). Este teste simplesmente soma a variância presente em cada dimensão (autovetor ou componente principal) e calcula o percentual em comparação a variância total dos dados. Ou seja, soma os autovalores dos componentes principais que não são ignorados e compara percentualmente com a soma de todos os autovalores, conforme eq. (2.4) [17],

$$PCV = \frac{\sum_{z=1}^u \gamma_z}{\sum_{z=1}^k \gamma_z} \times 100\% \quad (2.4)$$

onde,

γ_z : autovalor z correspondente à componente principal z usada;

u : número de componentes principais usadas;

k : número original de dimensões.

Se for determinado um percentual de variância mínimo que deve ser mantido da variância original, o teste mostrará quantas dimensões em termos de componentes principais devem ser utilizadas para os dados em análise.

2.5 Redes Neurais

O mecanismo natural de aprendizado, adaptação, reconhecimento e memória inspira muitos métodos matemáticos que buscam uma espécie de inteligência, a chamada artificial, que são métodos com a capacidade de aprender continuamente com dados de sistemas e que o reconhece efetivamente.

Um destes modelos ditos inteligentes é a Rede Neural, cujo próprio nome indica a inspiração biológica. Busca no funcionamento dos neurônios humanos uma maneira de gerar modelos matemáticos que aprendem com os dados informados e que se atualizam com treinamento [21].

A Rede Neural é utilizada quando se tem dados numéricos disponíveis. Outros métodos de Inteligência Artificial não necessariamente necessitam de entradas e saídas numéricas. A exemplo, o estudo *Fuzzy*, um método introduzido num artigo de 1965 por Lofti A. Zadeh. Este pode ser configurado a partir de conhecimentos linguísticos, possui saídas linguísticas, definida como variável linguística [22, 23].

As Redes Neurais possuem uma distinção interna entre algoritmos de aprendizados supervisionado e não-supervisionados. A diferença seria na necessidade ou não de um “professor”. Por exemplo, nas supervisionadas, as entradas e as saídas são fornecidas para treinamento da Rede Neural, ou seja, o modelo matemático a ser treinado se ajustaria tendo o conhecimento de onde partir (entradas do processo, x) e de para onde ir (saídas do processo, y). Em não supervisionadas, apenas as entradas do processo são fornecidas à Rede Neural e, sem ajuda de “professor” para ensinar a saída em seu treinamento [21].

Experimentos utilizando redes neurais foram usados em [24] utilizando redes supervisionadas para verificar se era possível a detecção a localização de vazamentos de sinais no domínio do tempo que eram filtrados para possuírem apenas 3 frequências em análise. As redes estudadas se mostraram capazes de realizar o diagnóstico.

2.5.1 Redes Neurais Supervisionadas

Ao se criar uma Rede Neural supervisionada para um sistema em análise, devem-se definir quais são as entradas do sistema (de modo que não sejam correlacionadas) e quais são as saídas do sistema, ou seja, as respostas finais desejadas e que a rede supervisionada deve gerar um modelo matemático capaz de calcular.

Deste modo, para o treinamento deste tipo de rede, é necessário ter conhecimento das saídas para que o modelo matemático funcione e convirja para os resultados esperados.

A saída desta rede é numérica e o treinamento é feito no sentido de diminuir o erro do modelo matemático gerado em relação aos valores de saída experimentais. Em geral, as redes neurais não devem ser utilizadas em casos de extrapolações visto que o modelo matemático gerado não tem qualquer informação fenomenológica do sistema. Assim, devem ser usadas apenas com dados correspondentes ao treinamento, ou em interpolações.

As redes neurais geradas, supervisionadas ou não, são específicas para a situação de treinamento. Se os valores das variáveis de saída (para rede supervisionada), ou os padrões encontrados nos dados de entrada (para rede não supervisionada), são dependentes de características do sistema, então alterações não treinadas pela rede, como mudança no material da tubulação, no material vazante, se a tubulação é subterrânea ou não, são extrapolações na resposta da rede.

Nos casos de alterações em, por exemplo, diâmetro de vazamento, porém compreendido entre o maior valor de diâmetro e o menor treinados, são considerados, neste trabalho, casos de interpolações.

2.5.2 Redes Neurais Não Supervisionadas

As redes não supervisionadas (do inglês, *Self-Organizing Maps*, SOM) são sistemas auto-organizáveis, ou auto-ajustáveis. Os treinamentos deste tipo de rede não possuem

necessariamente os valores que se desejam alcançar (y) no conjunto de dados de treinamento, apenas as entradas do sistema (x).

O propósito principal de um algoritmo não supervisionado é o de reconhecer padrões significativos em dados de entrada. Os algoritmos de treinamento de redes auto-organizáveis seguem regras em função da localidade dos dados de entrada [21].

Uma classe de aprendizados não-supervisionados é o de aprendizado competitivo. Nesta rede, neurônios competem entre si para serem atualizados no treinamento. Apenas um neurônio ganha a competição em relação a cada dado de entrada e a este se dá o nome de “neurônio vencedor”. A rede desenvolvida por Kohonen usa o aprendizado competitivo.

Em termos matemáticos, é declarado vencedor o neurônio calculado segundo a função discriminante de $j(x)$ [25]:

$$j(x) = \operatorname{argmin}_i \|x - w_i\| \quad (2.5)$$

onde x é qualquer valor de entrada e w_i são os vetores pesos. Nesta função, o neurônio, vetor peso, acusado pela função $j(x)$ é aquele com a menor distância Euclidiana ao vetor x . Ou seja, a função discriminante calcula o neurônio vencedor (c) em relação a este dado de entrada.

Neste trabalho, a nomenclatura “neurônio ativado” também será utilizada para designar os neurônios com as menores distâncias euclidianas em relação aos dados x em questão, seja de treinamento (para atualização dos vetores pesos durante treino) ou de teste (para avaliar a eficiência do mapeamento gerado). Testes de interpolações e de extrapolações podem ser avaliados, pois os dados sempre ativarão os neurônios com maiores similaridades, ou seja, os mais próximos (por distância Euclidiana).

2.5.2.1 Redes de Kohonen

Esta ferramenta é muito utilizada para visualização de dados multidimensionais posto que organiza informações diversas num espaço dimensional mais fácil de ser diretamente analisado.

Redes capazes de se auto-ajustar foram estudadas por Teuvo Kohonen e sua pesquisa pode ser encontrada no livro [26]. Na rede de Kohonen, o neurônio é representado por um vetor peso. O vetor peso, ou neurônio, possui a mesma dimensionalidade que os dados de entrada de treinamento.

A saída desta rede é um mapa de vetores pesos (ou neurônios) arranjado conforme o algoritmo de treinamento. Os neurônios, antes do treinamento, são inicializados e possuem conexões entre si arranjadas de forma que estes possuam vizinhos imediatos, com conexões de um ao outro [21]. A Figura 2.2 representa a organização da vizinhança, sem explicitar as conexões, entre neurônios: (a) em organização hexagonal e (b), retangular. A saída hexagonal é a estrutura mais comumente usada.

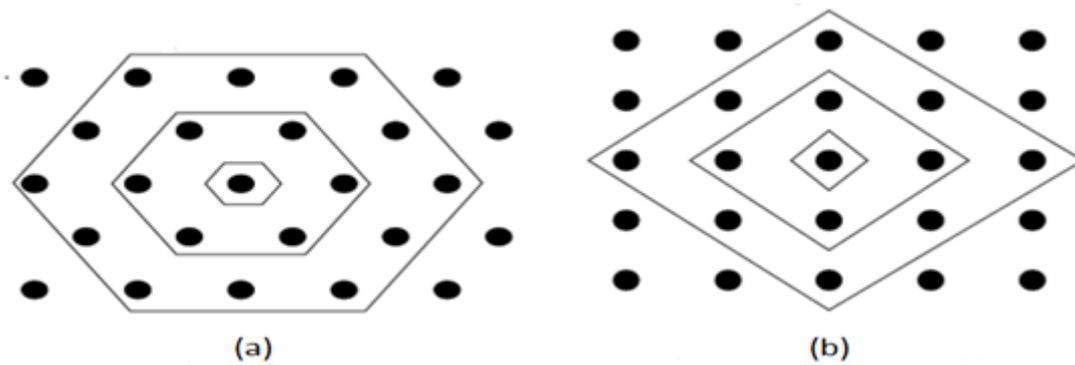


Figura 2.2 Esquema da organização de vizinhança dos neurônios na camada de saída. (a) hexagonal, (b) retangular. Adaptado de [27]

Percebe-se que na organização da vizinhança, cada neurônio possui seis vizinhos imediatos na organização hexagonal e, na retangular, quatro.

Na Figura 2.3, as vizinhanças de cada neurônio estão organizadas com cada neurônio possuindo quatro conexões a vizinhos imediatos (organização retangular), sem mostrar as conexões, porém em diferentes topologias da rede.

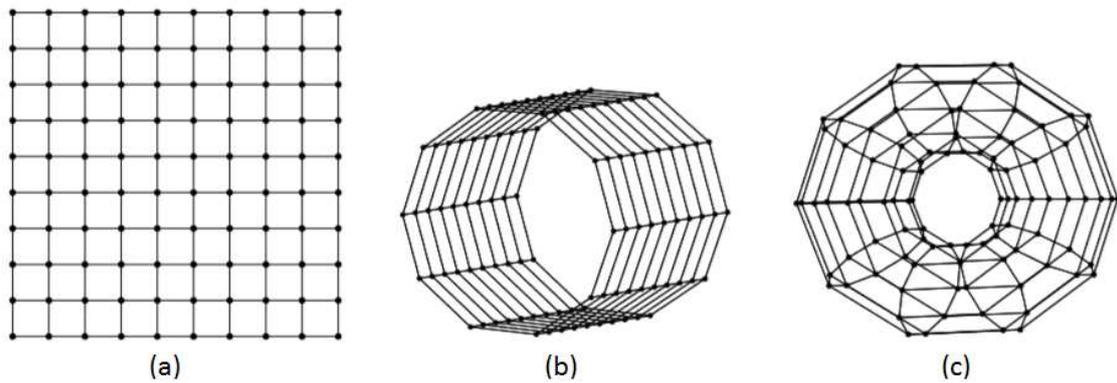


Figura 2.3 Topologias diferentes para o mapa auto-organizável de Kohonen: (a) formato plano; (b), cilíndrico e (c) toróide. Adaptado de [27]

Da Figura 2.3, as topologias (b) e (c) são normalmente propostas para dados de entrada circulares.

1) Inicializações da Rede

Os neurônios da rede de Kohonen são normalmente inicializados de forma randômica ou linear: a inicialização randômica tem valores aleatórios entre os valores máximos e mínimos de cada variável; a inicialização linear calcula primeiramente os autovalores e os autovetores da matriz de covariância dos dados de treinamento e os pesos são inicializados normalmente ao longo dos dois autovetores com maiores autovalores (componentes principais) calculados [27]. A Figura 2.4 mostra as duas inicializações em mapa com vinte e cinco neurônios no formato 5x5.

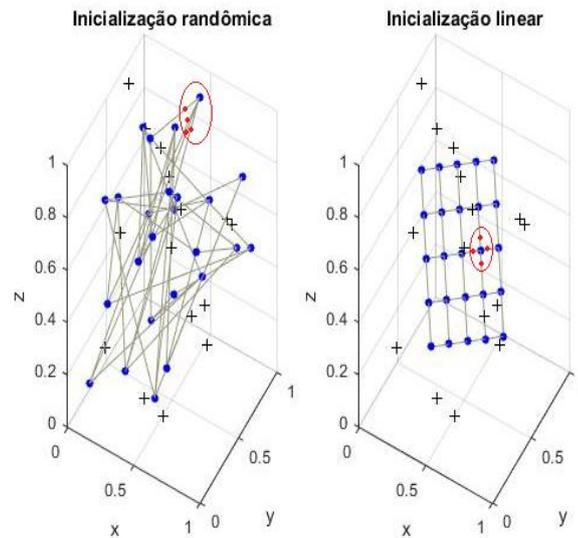


Figura 2.4 Inicializações dos vetores pesos, em azul. À esquerda, randômica; à direita, linear. Dados de treino randômicos simbolizados por “+”. Gerados no *software Matlab*

Os vetores de peso em inicializações lineares têm sempre as mesmas coordenadas para um determinado dado de treinamento, enquanto que, na inicialização randômica, modificam de treinamento para treinamento.

Percebe-se que, na Figura 2.4, a organização da vizinhança é retangular, o que faz com que cada neurônio (que não esteja nas extremidades) tenha quatro conexões com vizinhos, chamados imediatos. Em ambas as inicializações, um neurônio teve suas quatro conexões evidenciadas em vermelho. Após inicialização, ocorre o treinamento.

As conexões dos neurônios com seus vizinhos imediatos são estabelecidas na inicialização e não se alteram após treinamento. Por mais que, possivelmente, estes neurônios se afastem, ou não sejam os mais próximos em distâncias Euclidianas, as conexões ainda fazem deles “imediatos”.

2) Treinamentos da Rede

Dois algoritmos de treinamento da rede auto-organizável são usados neste trabalho: algoritmo sequencial e algoritmo em batelada.

O objetivo do treinamento é o de atualizar os vetores pesos inicializados de modo que estes reconheçam padrões nos dados de entrada. Neste trabalho, os neurônios ditos treinados

são os vetores pesos resultantes das atualizações feitas pelo algoritmo de treinamento utilizado.

2.1) Algoritmo Sequencial

O algoritmo sequencial é um aprendizado competitivo no qual apenas o neurônio com a menor distância Euclidiana, acusado pela equação discriminante (2.5), em relação ao dado da entrada (x) é ativado. No treinamento, o neurônio, que recebe o nome de “vencedor”, é então atualizado. O efeito da atualização do vetor peso faz com que este se aproxime ainda mais do respectivo dado de entrada que o ativou, representando-o ainda mais. A Figura 2.5 representa um dado de entrada computando o neurônio com a menor distância. Deste modo, conforme o treinamento progride, cada neurônio fica mais fortemente associado àquelas entradas que o ativam, ou a entradas com padrões próximos.

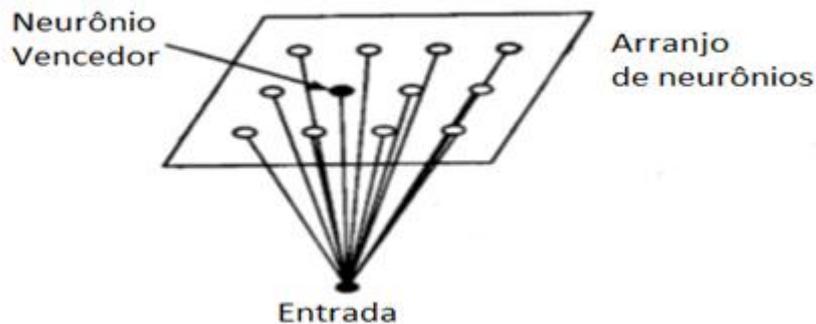


Figura 2.5 Mapa de Rede Auto-Organizável de Kohonen. Adaptado de [25]

A atualização de pesos do neurônio pelo algoritmo sequencial está representada na equação (2.6) [28]. O nome é sequencial pois processa um vetor de entrada, x , por vez.

$$w_i(t + 1) = w_i(t) + \alpha(t)h_{ic}(x - w_i(t)) \quad (2.6)$$

Na eq. (2.6), x é o vetor de entrada que elegeu o respectivo neurônio vencedor; $w_i(t + 1)$ representa o vetor peso atualizado de $w_i(t)$ peso inicial de qualquer neurônio que venha a ser atualizado (vencedor e sua vizinhança); $\alpha(t)$ representa a taxa de aprendizado;

h_{ic} , a função de vizinhança ao redor do neurônio vencedor (c) e t é o número da iterações. A atualização ocorre até que o número máximo de iterações estabelecido seja alcançado.

Quando todas as observações de entrada são processadas, todos os diferentes valores de x , e atualizados os respectivos pesos, diz-se que uma época (ou iteração) se completou [29], conceito também usado o treinamento de redes supervisionadas. A ordem das amostragens dos dados de treino, um valor x por vez, é aleatória, assim, o algoritmo sequencial pode gerar resultados diferentes para um mesmo dado de treino.

Há algumas funções diferentes que definem a taxa de aprendizado, $\alpha(t)$. A melhor definição a ser usada varia de caso para caso. Porém, de acordo com [21], a taxa de aprendizado deve inicialmente, nas primeiras 1.000 iterações, possuir valor próximo a 1 enquanto diminui conforme o número de iterações aumenta, porém se mantendo sempre acima de 0,1. Nesta fase inicial, ocorre a ordenação topológica dos vetores de peso, caracterizando a fase de ordenação. No restante do treinamento, a fase de refinamento, o valor de $\alpha(t)$ já atinge grandezas na ordem de 0,01 ou menos depois de muitas iterações.

Algumas funções utilizadas para a taxa de aprendizado são a “inversa”, “power” e “linear” definidas pelas equações a seguir [27].

$$\alpha(t)_{inversa} = \frac{\alpha_0(t)}{1 + \left(1 + \frac{100t}{T}\right)} \quad (2.7.1)$$

$$\alpha(t)_{power} = \alpha_0(t) \left(\frac{0,005}{\alpha_0(t)}\right)^{t/T} \quad (2.7.2)$$

$$\alpha(t)_{linear} = \alpha_0(t) \left(1 - \frac{100t}{T}\right) \quad (2.7.3)$$

Com $\alpha_0(t)$ sendo a taxa inicial de aprendizagem, t a iteração atual e T o número máximo de iterações. Um valor de $\alpha_0(t)$ usual é de 0,5 para a fase de ordenação topológica e de 0,01 na fase de refinamento [30].

De maneira semelhante, a função de vizinhança, h_{ic} , também possui diferentes formatos e também é decrescente com o aumento de iterações [28]. Sua escolha afeta na organização topológica do mapa gerado.

Quando o neurônio vencedor é atualizado pelo algoritmo, sua vizinhança pode ser atualizada também. A função de vizinhança, h_{ic} , define a intensidade desta atualização da vizinhança. Algumas definições de função de vizinhança estão a seguir: “*bubble*”, “*gaussian*” e “*cutgauss*” [27].

$$h_{ic_{bubble}} = 1(\sigma_t - d_{ci}) \quad (2.8.1)$$

$$h_{ic_{gaussian}} = e^{-d_{ci}^2/2\sigma_t^2} \quad (2.8.2)$$

$$h_{ic_{cutgauss}} = e^{-d_{ci}^2/2\sigma_t^2} (\sigma_t - d_{ci}) \quad (2.8.3)$$

onde, $(\sigma_t - d_{ci})$ é uma função *step* em que: se $\sigma_t \geq d_{ci}$, $(\sigma_t - d_{ci}) = 1$; se $\sigma_t < d_{ci}$, $(\sigma_t - d_{ci}) = 0$.

σ_t : é o raio da vizinhança em relação ao neurônio vencedor. Este decresce com o número de iterações; como raio inicial, pode-se usar, por exemplo, 2. Este valor decresce por exemplo, linearmente com o aumento do número de iterações, até o valor final do raio, por exemplo, 1.

d_{ci} : é a distância euclidiana do neurônio vencedor (c) ao neurônio i .

Das definições apresentadas, valor máximo de h_{ic} é a unidade. Para o neurônio vencedor, o valor de $h_{ic} = 1$ ocorre nas três definições de função de vizinhança acima.

Nas definições “*bubble*” e “*cutgauss*” se o neurônio se encontrar dentro do raio da vizinhança (σ_t), tem seu valor atualizado pelo algoritmo com o valor de h_{ic} calculado pela respectiva definição. Na definição “*gaussian*”, não há limites espaciais para a atualização do vetor peso da vizinhança, porém, quanto mais distante, menor o valor de h_{ic} calculado.

Em geral, nas primeiras 1.000 iterações, de ordenação topológica, o raio da função de vizinhança em relação ao neurônio vencedor diminui até que apenas o neurônio vencedor mais os vizinhos próximos sejam atualizados com vetor de entrada x . Neste ponto, deve-se começar a fase de refinamento [21]. A Figura 2.6 ilustra a progressão da ordenação dos neurônios conforme o número de iterações aumenta.

As escolhas da taxa de aprendizado ($\alpha(t)$) e da função de vizinhança (h_{ic}) afetam a resposta final da rede, suas escolhas são normalmente feitas por tentativas e erros, avaliando a melhor configuração da resposta final [21].

2.2) Algoritmo em Batelada

O algoritmo em batelada é um treinamento também iterativo. Em vez de computar uma entrada x por vez, todos os dados de treinamento são inicialmente apresentados ao algoritmo. Assim, os dados de entrada são primeiramente divididos de acordo com as regiões de *Voronoi* dos vetores pesos inicializados [27]: cada região compreende um vetor peso (neurônio) e todos os dados que se encontram mais próximos a este neurônio do que a qualquer outro. Ou seja, os dados compreendidos na região *Voronoi* de um neurônio são todos os dados de treino que elegeriam o neurônio como “vencedor”.

Assim, o algoritmo de treinamento do vetor peso em batelada está a seguir [27].

$$w_i(t + 1) = \frac{\sum_{j=1}^n h_{ic}(t)x_j}{\sum_{j=1}^n h_{ic}(t)} \quad (2.9)$$

Assim, mais de uma entrada x é utilizada por vez para atualização dos pesos. O valor n é o número de dados de treino na região *Voronoi* do respectivo vetor peso. Neste algoritmo, não se utiliza a função $\alpha(t)$. Como as regiões *Voronoi* de um mesmo dado de entrada são sempre as mesmas, esse algoritmo, por si, dá as mesmas respostas para um mesmo treinamento.

Se forem combinadas inicializações lineares dos pesos com treinamento com o algoritmo em batelada, sempre haverá a mesma resposta final de rede para um mesmo dado de entrada.

3) Mapeamento dos Neurônios: Erros de Quantização e Topográfico

Como mencionado anteriormente, quando o vetor peso é atualizado, ele se aproxima ainda mais do vetor de entrada x que o ativou (ou, em batelada, vetores de entrada). Assim, na progressão do treinamento, mais fortemente os neurônios representam os vetores de entrada. A Figura 2.6, mostra a progressão dos vetores peso esperada com aumento de iterações em um espaço com dados de treinamento (não representados).

Dois parâmetros são normalmente usados para se determinar a qualidade de um treinamento: o erro quantização e o erro de topográfico. O erro de quantização representa quão bem o vetor de entrada é representado por seu respectivo neurônio vencedor: é a média

das distâncias de cada vetor de entrada em relação ao seu neurônio mais próximo, o vencedor [28]. Enquanto que, o erro topográfico é a quantidade média de vezes em que o segundo neurônio mais próximo de uma entrada x não é vizinho imediato do primeiro [29].

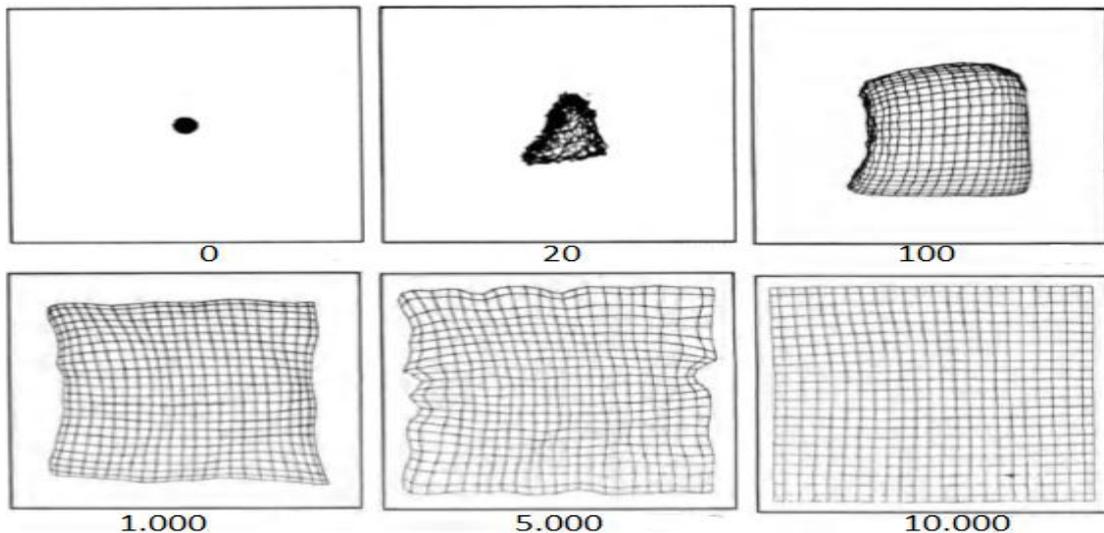


Figura 2.6 Ordenação dos pesos com o aumento de iterações de treinamento. Adaptado de [26]

Ao final do treinamento, espera-se que neurônios próximos um ao outro representem ou sejam ativados por- vetores de entrada x com mesmos padrões, ou padrões próximos. O Apêndice A mostra um exemplo com o passo-a-passo do treinamento de uma rede de Kohonen.

2.5.2.2 Matrix U

Os neurônios são vetores com mesma dimensão dos vetores de entrada. Em dados multidimensionais como entrada, um gráfico dos neurônios é tão limitado quanto dos dados de entrada. Para se visualizar os neurônios treinados, utiliza-se comumente a matriz U (em inglês *unified distance matrix, u-matrix*). A visualização da rede treinada a partir da matriz U possibilita o estudo de padrões com alta dimensionalidade em poucas dimensões, comumente, duas dimensões apenas.

Esta matriz possui formato plano em que colorações representam as distâncias Euclidianas dos neurônios com seus vizinhos imediatos. Normalmente, tonalidades claras indicam proximidade e escuras, distanciamento.

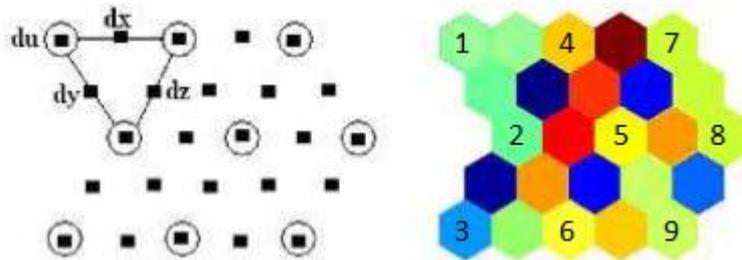


Figura 2.7 Esquema de geração da matriz U em vizinhança hexagonal de neurônios 3x3.

Adaptado de [31]

Neurônios próximos um do outro estão associados a dados de entrada com mesmo padrão, ou padrões similares. A representação na matriz U dessas proximidades são áreas adjacentes de tonalidade clara.

Na Figura 2.7, a matriz U de um mapa auto-organizável com 9 neurônios, organizada em 3 linhas e 3 colunas- 3x3, em organização hexagonal, é esquematizada à direita. Os hexágonos contados de 1 a 9 representam os neurônios de mesma numeração, cuja coloração representa a mediana (média, máxima ou mínima [32]) da distância deste neurônio aos neurônios vizinhos imediatos. Os demais hexágonos, os sem numeração, possuem a coloração representando diretamente a distância entre os neurônios enumerados com quem possui fronteiras na matriz: por exemplo, o hexágono sem numeração situado entre o “1” e o “4” ilustra a distância entre esses dois neurônios. Os valores numéricos são substituídos por cores que variam de escuro (valores altos) a claro (valores baixos).

Desta forma, a matriz-U possibilita a investigação da topologia do mapa, acusando grupos semelhantes por regiões claras e fronteiras por regiões escuras.

Com o treinamento já realizado, a rede pode ser investigada de modo a saber quais neurônios representam quais padrões dos vetores de entrada x . Os vetores de entrada (sejam os utilizados no próprio treinamento quanto valores separados para teste) ativam o neurônio mais próximo, por distância Euclidiana. Espera-se que dados de mesmo padrão ativem os mesmos neurônios, ou neurônios próximos. Esta investigação também pode ser representada

na matriz U. Um exemplo se encontra também no Apêndice A, juntamente com a construção mais detalhada da matriz U do exemplo.

2.5.3 Conclusões sobre Redes Neurais para a Proposta de Pesquisa

A rede não supervisionada de Kohonen é escolhida neste trabalho pela falta de necessidade de se atribuir valores de saída aos dados de entrada de treinamento da rede. Dessa forma, por mais complexa que seja a situação de anormalidade em ocorrência no sistema, a partir de cada padrão acústico, os neurônios serão mapeados de forma a diferenciar um padrão de outro.

A rede supervisionada, por outro lado, precisaria necessariamente de valores numéricos associados aos dados de entrada do sistema. Por exemplo, se se associasse o valor de saída como 0 a todas as situações sem ocorrência de vazamento, porém se se ocorressem perturbações como batidas ao longo da aquisição de dados, a rede poderia ter dificuldades em continuar calculando o valor 0. Além, no sistema de simulação de vazamentos deste trabalho, esta associação de valores numéricos se torna subjetiva e pode não convir. Enquanto que, na rede de Kohonen, espera-se que, contando que haja padrões diferentes, estes podem ser mapeados e distinguidos de modo a se identificar e caracterizar o sistema.

Neste trabalho, além de se criar um mapa com neurônios treinados por dados experimentais de diferentes configurações de vazamento, investigam-se quais neurônios são ativados por cada padrão diferente. O treinamento da rede é não-supervisionado, porém os testes serão supervisionados para a avaliação dos neurônios.

Espera-se, então, que as redes criadas se tornem aptas a identificar a perturbação do sistema de acordo com os neurônios ativados em qualquer teste futuro de monitoramento da tubulação.

Assim, propõe-se a utilizar do método acústico para obtenção de dados de amplitudes sonoras no domínio do tempo e no da frequência para análises e possíveis identificações do sistema. O domínio do tempo será analisado de forma direta como um reconhecimento mais superficial dos eventos em ocorrência no sistema e, possivelmente, gerar noções de localização do vazamento pela análise de amplitude da onda sonora captada na sequência dos microfones instalados. As demais investigações são com os dados no domínio da frequência, com redução da dimensionalidade, nas quais redes não supervisionadas de Kohonen são

treinadas para se obter formas de fazer uma análise do sistema mais completa sobre os eventos ocorridos no sistema como a dimensão do vazamento e sua localização aproximada.

3 METODOLOGIA

Com a finalidade de investigar a eficiência do método acústico em gerar dados suficientes para a rede neural de Kohonen detectar a ocorrência de vazamento e sua possível localização, experimentos foram realizados em tubulação laboratorial simulando situações de vazamento de gás e de outras perturbações, como batidas, que o sistema pode também estar sujeito na prática. Ao longo deste trabalho, a tubulação utilizada esteve montada no Laboratório de Controle e Automação de Processos (LCAP) da FEQ/Unicamp.

3.1 Sistema montado de Simulação dos Vazamentos

O sistema montado no laboratório para a simulação dos vazamentos consiste em regulador de pressão, botijão de armazenamento e tubulação de cobre.

O regulador de pressão tem a função de acomodar a pressão experimental àquela de interesse no início de cada simulação. O ar comprimido é proveniente da linha de alimentação aos laboratórios da faculdade e sua alimentação será mantida contínua durante todos os experimentos.

A pressão nas simulações será inicialmente ajustada para 1 kgf/cm^2 no regulador de pressão. Esta pressão está baseada nas tubulações de distribuição de gás de cozinha em instalações domésticas.

O botijão de armazenamento consiste em um vaso de pressão de aço carbono com capacidade de 34,5 l [5]. Ele será utilizado apenas para simulações de batidas em seu exterior.

E, a tubulação metálica de cobre é de aproximadamente 53 m e 1/2" de diâmetro. A Figura 3.1 mostra a tubulação dos ensaios.



Figura 3.1 Sistema utilizado para simulações de vazamentos (LCAP/FEQ/Unicamp)

A entrada do ar na tubulação está circulada na Figura 3.1 onde indicado “ar”, o botijão em “Botijão de Armazenamento” e o regulador de pressão, “reg. de p.”. Assim, o caminho que o ar comprimido percorre, ao sair da linha de alimentação da faculdade é na ordem: regulador de pressão, botijão de armazenamento e, por fim, a tubulação.

Na tubulação, há 7 furos realizados por broca em posições e tamanhos de diâmetro diferentes. Esses furos são utilizados para simular os vazamentos. Quando não utilizados, estão cobertos por fita isolante evitando fuga de ar comprimido.

As posições dos furos e suas magnitudes estão na Tabela 3.1.

Tabela 3.1 Distâncias dos furos com relação a entrada do gás comprimido na tubulação

| Furo | Distância | Diâmetro | Furo | Distância | Diâmetro |
|--------------|-----------|----------|------------|-----------|----------|
| Primeiro, F1 | 0,2 m | 1,0 mm | Quinto, F5 | 3,4 m | 4,0 mm |
| Segundo, F2 | 1,0 m | 0,5 mm | Sexto, F6 | 4,8 m | 3,0 mm |
| Terceiro, F3 | 1,6 m | 2,0 mm | Sétimo, F7 | 5,6 m | 0,5 mm |
| Quarto, F4 | 2,4 m | 1,0 mm | - | - | - |

3.2 Dispositivos para aquisição

A aquisição é formada por um conjunto de equipamentos. Estes são utilizados em conjunto de forma a captar, transmitir e processar a informação do sistema em estudo com simulações de vazamento.

3.2.1 Microfones

Microfones são transdutores capazes de converter energia sonora em elétrica. Acoplados à estrutura metálica, encontram-se 15 microfones, mas apenas 5 são utilizados no experimento por limitações no processamento de informações.

Em situação experimental, as ondas acústicas capturadas pelos microfones são o foco. Os microfones são do modelo XCM-9767 omnidirecionais. A curva da resposta de frequência deste modelo está representada pela Figura 3.2.

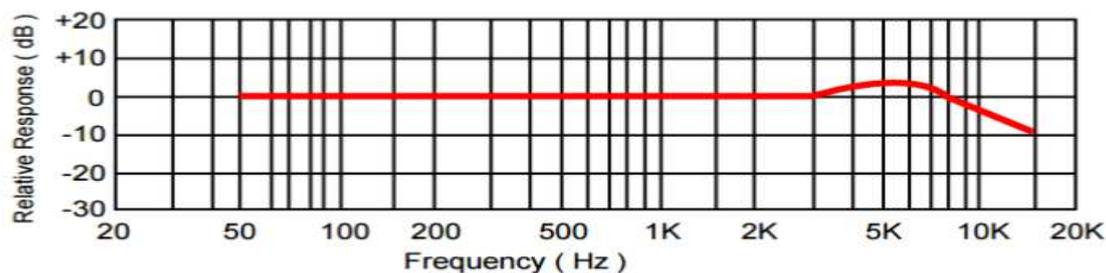


Figura 3.2 Resposta de Frequência do microfone modelo XCM-9767 [5]

Em curva de frequência como a da Figura 3.2, as melhores faixas para se trabalhar com o microfone são aquelas em que sua resposta relativa seja unitária, ou nula se representada em decibéis, não interferindo no sinal de saída. A capacidade deste microfone é de 50 Hz à 15.000 Hz, aproximadamente. Porém, o intervalo de frequência em estudo é de 100 Hz à 10.000 Hz, com base na tese [5] que estudou de 500 à 16.000 Hz, apesar de apresentar resposta relativa diferente deste valor em frequências acima de 3.000 Hz.

As posições dos microfones estão na Tabela 3.2. Os 5 microfones, escolhidos de forma a obter observações ao longo de toda a tubulação nos experimentos realizados neste trabalho são: m1, m2, m6, m11 e m13.

Tabela 3.2 Distância dos microfones com relação a entrada do gás comprimido na tubulação

| Microfone | Distância | Microfone | Distância |
|-----------------|-----------|-------------------|-----------|
| Microfone 1, m1 | 0,1 m | Microfone 9, m9 | 31,0 m |
| Microfone 2, m2 | 3,3 m | Microfone 10, m10 | 34,8 m |
| Microfone 3, m3 | 7,7 m | Microfone 11, m11 | 38,6 m |
| Microfone 4, m4 | 11,4 m | Microfone 12, m12 | 42,5 m |
| Microfone 5, m5 | 15,5 m | Microfone 13, m13 | 45,7 m |
| Microfone 6, m6 | 19,2 m | Microfone 14, m14 | 49,3 m |
| Microfone 7, m7 | 23,2 m | Microfone 15, m15 | 54,4 m |
| Microfone 8, m8 | 27,1 m | - | - |

Avaliando a Tabela 3.1 juntamente com a Tabela 3.2, todos os furos se encontram entre o microfone 1 e o microfone 3.

3.2.2 Pré-amplificador

O pré-amplificador é utilizado para diminuição da impedância do microfone. Assim, menor a perda de energia na transmissão do sinal elétrico através de cabos [6].

O circuito utilizado foi desenvolvido na dissertação [6] e foi originalmente projetado para acomodar apenas um microfone. Segundo a dissertação, o circuito desenvolvido possui três estágios: no primeiro é feita a conversão de impedância na intenção de garantir a máxima transferência de sinal do microfone para o pré-amplificador de tensão, no segundo e no terceiro são realizadas amplificações do sinal com a utilização de amplificadores operacionais.

No presente trabalho, o circuito do pré-amplificador encontra-se modificado para acomodar 15 microfones.

3.2.3 Conversor Analógico/Digital e CPU

Após o amplificador, o sinal analógico é convertido em digital utilizando o conversor A/D NI cDAQ-9178. O sinal digital, então, é processado por um computador contendo o software *Labview*.

Assim, os dispositivos de aquisição de dados são o conjunto de microfones, pré-amplificador, conversor analógico-digital e o computador com o programa de aquisição de dados *Labview*.

3.3 Aquisição de Dados no Domínio do Tempo e da Frequência

O software *Labview* é utilizado para a aquisição de dados no domínio do tempo e cálculo do domínio da frequência.

O ambiente virtual do *Labview* (a página de nome “*Front Panel*”) é um espaço virtual onde se configuram os comandos que se queiram (seja de aquisição, de cálculos etc) realizados no *software*: blocos virtuais são posicionados de acordo com suas funções e na ordem em que se desejam. Este ambiente, conforme será utilizado neste trabalho, está representado na Figura 3.3. A linha azul representa o caminho da informação que é adquirida pelo primeiro bloco, o de nome “*DAQ Assistant*”.

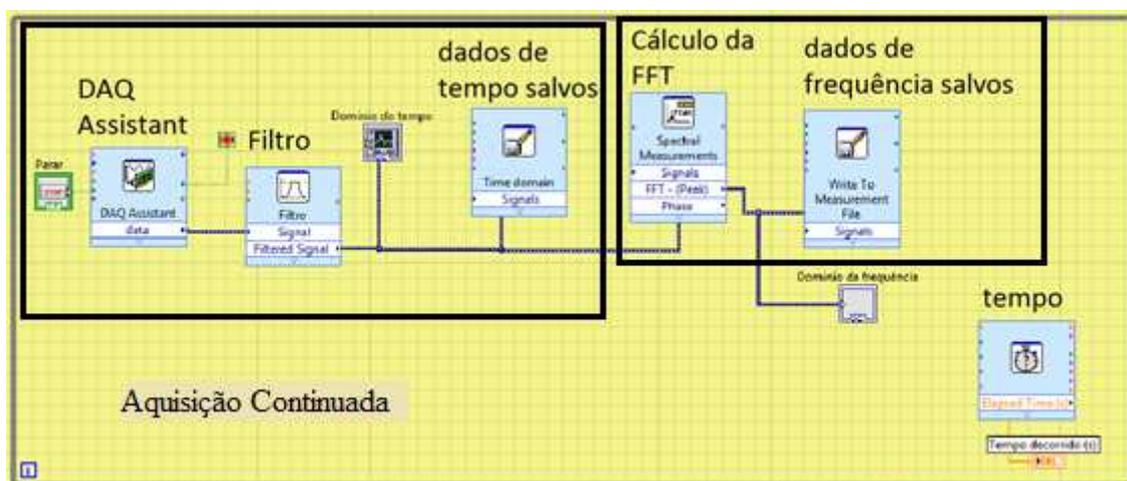


Figura 3.3 Ambiente virtual do Labview para aquisição de dados, página “*Front Panel*”

A aquisição de dados será realizada continuamente assim como o cálculo do domínio da frequência. Na Figura 3.3, todo o sistema virtual está compreendido em um retângulo maior de coloração cinza, este é um *loop* similar ao comando *while* comuns em programações C ou C++: tudo no interior do *loop* é repetido até que a uma condição se atinja. A condição deste programa é a ativação do botão virtual “parar” e este foi criado para ser ativado no momento em que se deseja encerrar a aquisição. Neste trabalho, a aquisição terá duração de 60 s.

3.3.1 Aquisição dos dados no domínio do tempo

A aquisição de dados no domínio do tempo terá três etapas: definição de parâmetros para a aquisição, filtragem de frequências não desejadas e armazenar os dados resultantes. Estas etapas estão evidenciadas pelo retângulo à esquerda da Figura 3.3 e expostas nesta seção.

1) Definição de Valores de Parâmetros Para a Aquisição

No ambiente virtual do *Labview*, determinam-se parâmetros de aquisição de dados em um bloco virtual chamado *DAQ Assistant Express*: o sinal contínuo amplificado do microfone, na faixa compreendida entre -10 e 10V, é adquirido pelo sistema digital de monitoramento numa frequência de amostragem de 30.000 Hz (o suficiente para evitar o fenômeno *aliasing*, acima de duas vezes 10.000 Hz, a frequência mais alta em análise).

A transferência de dados amostrados do *DAQ* para a memória do computador é definida em “*Samples to Read*” e será ajustada em 20.000. Assim, os dados serão guardados de 20.000 em 20.000 pontos.

2) Filtro

No ambiente virtual de aquisição de dados do *Labview*, outro bloco também será utilizado antes dos dados no tempo serem armazenados: um filtro virtual. Este filtra valores abaixo de 100 Hz e acima de 10.000 Hz através de filtro passa-faixa de topologia *Butterworth*, a escolha da faixa de frequência teve como base a tese [5] com a faixa de estudo de 500 Hz a 16.000 Hz, onde as frequências acima de 10.000 Hz não foram muito utilizadas e as mais baixas, próximas de 500 Hz, mostraram em geral grande atividade.

O filtro de *Butterworth* possui a característica de possuir uma saída tão plana (sem atenuação) quanto matematicamente possível na faixa passante e de se aproximar do filtro ideal quando sua ordem aumenta. A ordem escolhida será 10 por ser a maior ordem disponível no *Labview*.

3) Armazenamento de Dados

O Bloco com os dizeres “Dados de tempo salvos” é responsável armazenamento dos dados em aquisição na memória do computador enquanto o experimento está em execução.

Assim, a aquisição de dados no domínio do tempo no *Labview* possui as etapas:

1. Aquisição de dados dos microfones (amplificada e digitalizada) no domínio do tempo;
2. Filtragem da onda capturada mantendo a faixa de passagem entre 100 Hz e 10.000 Hz;
3. Valores armazenados (dados no domínio do tempo);

Fim da aquisição de dados.

3.3.2 Cálculo das Amplitudes no Domínio da Frequência

Após a filtragem, a informação capturada e armazenada, que se encontra no domínio do tempo, passará por um bloco chamado “*Spectral Measurements*” onde a Transformada Rápida de Fourier é realizada continuamente ao longo da aquisição dos dados. Este Bloco, também presente na Figura 3.3 nos dizeres “Cálculo da FFT”, será configurado para “*Magnitude (Peak)*” no qual a magnitude do sinal é medida em termos de seu valor máximo, a amplitude da onda na frequência em análise. O resultado será “linear”, que conserva a saída em sua unidade original (contra a opção “dB” que transformaria a unidade de saída em decibéis). Neste mesmo bloco, também se acrescentará ao sinal uma função *Window* para se compensar descontinuidades do sinal, principalmente por ser função do tempo.

Logo após o cálculo do domínio da frequência, os dados calculados também serão armazenados. Estes blocos estão evidenciados na Figura 3.3 pelo retângulo à direita.

3.4 Redução da Dimensionalidade e Redes Neurais

Em seguida, os resultados obtidos no domínio do tempo e no domínio da frequência serão exportados ao *software Matlab* onde os resultados de identificação e de caracterização e localização dos vazamentos serão obtidos.

3.4.1 Investigações com Dados no Tempo

Os dados no domínio do tempo serão investigados como uma forma direta e rápida de se obter o comportamento do sistema pela análise do gráfico direto das amplitudes no domínio do tempo armazenadas pelo *Labview* contra o respectivo tempo.

Além disso, as variações das amplitudes médias, causadas pela ocorrência de perturbações, de dados de cada microfone utilizado, serão calculadas para avaliar quais perturbações mais modificam o sistema. Além disso, investigar possíveis correlações das variações com a localização do vazamento em relação ao respectivo microfone.

3.4.2 Investigações com Dados no Domínio da Frequência

O domínio da frequência será a entrada da rede de Kohonen e terá seus padrões mapeados por neurônios no treinamento não-supervisionado. Mas, antes do treinamento, a dimensionalidade dos dados em frequência deverá ser reduzida para evitar grandes esforços computacionais desnecessários (com dados de informações redundantes, ou insignificantes).

3.4.3 Redução da Dimensionalidade dos Dados em Frequência

Os dados adquiridos no domínio da frequência serão analisados em suas componentes principais para diminuir a dimensão dos dados. As variáveis de entrada para o PCA são as observações de amplitude de onda em respectiva frequência (varrendo o intervalo de 100-10.000 Hz) calculadas pelo FFT. Ou seja, com os dados em termos de frequências, cada amplitude de onda na respectiva frequência é uma dimensão. Assim, as observações estão num espaço multidimensional, posto que há observações ao longo do tempo de aquisição de dados.

O programa para realização do PCA possui as etapas:

1. Leitura dos dados no domínio da frequência;
2. Padronização dos dados (média de cada dimensão igual a 0 e desvio padrão igual a 1);
3. Uso do algoritmo do *Matlab* para cálculo direto dos autovalores e autovetores da matriz de covariância dos dados. Este algoritmo também gera os dados já em termos de suas componentes principais;

4. Escolha de quantidade de componentes principais (dimensões) necessárias para manter boa representatividade dos dados originais com o teste percentual cumulativo da variância;
5. Geração de dados em dimensão menor em termos de componentes principais;
6. Fim do programa.

3.4.4 Criação de Redes para a Identificação, Caracterização e Localização de Vazamentos

Redes neurais serão geradas com dados comprimidos por PCA e, para investigar a possibilidade e a eficiência de outros tratamentos de dados, redes também serão geradas por dados diretamente do domínio da frequência.

Para a implementação dos comandos utilizados para treinamento e teste da rede não supervisionada de Kohonen no *Matlab*, o *toolbox* disponível na internet [33] sob o nome de *Self-Organizing Map Toolbox* (SOM) foi utilizado.

1) Redes Treinadas com Dados em Dimensão Reduzida por PCA

Uma vez que os dados se encontrem em dimensão reduzida, eles servem de entrada para as redes auto-organizáveis de Kohonen criadas no *Matlab* para, finalmente, ocorrer a determinação de ocorrência, ou não, de vazamentos. Dois objetivos serão almeçados: determinar localização e determinar dimensão do vazamento.

1.1) Criação da Rede Neural

A rede com o objetivo de determinar a localização terá dados dos cinco microfones utilizados na aquisição; enquanto que, as redes com o objetivo de determinar a dimensão do vazamento (se há batidas, ou evento normal), terá em seus dados de treinamento apenas dados de um microfone.

A criação das redes não supervisionadas de Kohonen seguem a ordem:

1. Leitura dos dados oriundos da análise do PCA;
2. Normalização dos dados;
3. Separação (se houver) de dados apenas de um microfone;
4. Separação de dados para treino e para teste;

5. Criação da rede de base (comando *som_make* do *toolbox* supracitado, usa equações pré-definidas para construção da rede);
6. A partir da rede de base, testar melhores configurações de quantidade de neurônios, topologia, algoritmo de treinamento e demais parâmetros da rede (resultados comparados pelos erros e pela configuração da matriz U gerada);
7. Treinar a rede, com a configuração escolhida na etapa 6, com dados de treinamento;
8. Computar erro de quantização e topográfico;
9. Obtenção do mapa de neurônios coordenados e matriz U;
10. Verificação dos neurônios ativados pelos dados de treino na matriz U;
11. Verificação dos neurônios ativados com dados de teste na matriz U;
12. Conforme necessidade, voltar a etapa 6 para redefinir configuração da rede;
13. Fim do programa.

Nas etapas 10 e 11, os neurônios são evidenciados conforme o objetivo da rede: se o objetivo da rede é determinar a localização, os neurônios serão evidenciados conforme a identificação de dados oriundos de microfones diferentes, ou seja, distâncias diferentes (independentemente da dimensão do vazamento, ou outras anormalidades simuladas).

Para as redes com o objetivo de determinar o evento que ocorre no sistema, nas etapas 10 e 11, os neurônios serão evidenciados de acordo com quais são ativados por dados de dimensões diferentes de vazamentos, ou de ausência de vazamentos.

O Apêndice A mostra a programação de uma das redes criadas, com dados do microfone 6 apenas, para identificação da dimensão dos vazamentos, batidas, ou ausência de anormalidades.

1.2) Teste de Interpolação

Um teste de interpolação foi feito em uma rede treinada sem dados de vazamentos de diâmetro intermediário (F3 de 2 mm), usando-os apenas como dados de teste.

As etapas da criação da rede para teste de interpolação são:

1. Leitura dos dados oriundos da análise do PCA;
2. Normalização dos dados;

3. Separação de dados apenas do primeiro microfone, m1 (escolha aleatória do microfone);
4. Separação de dados para treino (excluindo totalmente dados que incluam vazamentos em F3) e para teste (contendo dados incluindo vazamentos em F3);
Após estes quatro passos, seguem-se os passos 5 a 13 da subseção 1.1 anterior.

1.3) Testes de Extrapolações

Para avaliar a eficiência da rede não supervisionada em casos de extrapolações, dois testes serão feitos:

- 1) Extrapolação por diâmetro de vazamento: em outra rede treinada sem dados de vazamento no maior diâmetro (F5 de 4 mm). Estes dados serão usados apenas como dados de teste.

As etapas para a criação da rede para teste de extrapolação por diâmetro de vazamento seguem a ordem:

1. Leitura dos dados oriundos da análise do PCA;
2. Normalização dos dados;
3. Separação de dados apenas do primeiro microfone, m1 (escolha aleatória);
4. Separação de dados para treino (excluindo totalmente dados que incluam vazamentos em F5) e para teste (contendo dados incluindo vazamentos em F5);

Após estes quatro passos, seguem-se os passos 5 a 13 da subseção 1.1 anterior.

- 2) Extrapolação em relação à distância dos vazamentos ao microfone: testes de dados capturados pelo segundo microfone, m2, testados em rede treinada com dados capturados m6.

Os comandos no *software Matlab* utilizados para a extrapolação em relação à distância estão na parte final do Apêndice B.

2) Redes Treinadas com Dados diretamente no Domínio da Frequência

Alguns testes serão realizados sem a compressão de dados por PCA. Meios de diminuir a dimensão ainda se tornam necessários para que apenas dados significativos estejam presentes no treinamento da rede e não tornem o esforço computacional excessivo sem necessidade.

No primeiro teste sem compressão por PCA, a seguir, a redução da dimensionalidade será realizada por uma amostra, escolhida aleatoriamente, de frequências para avaliação do mapa treinado pela rede não supervisionada diretamente com dados no domínio da frequência. Este teste tem o objetivo, se a escolha aleatória conseguir reunir dados representativos, de avaliar a influência -positiva ou negativa- de se comprimir os dados pelo PCA.

O segundo teste, verificará se dados no regime permanente são suficientes para diferenciar diâmetros de vazamento. Assim, apenas uma pequena amostra dos dados, contando que se esteja em regime permanente, seria necessária.

2.1) Redução da Dimensionalidade por Exclusão de Frequências na criação da Rede

Os primeiros testes sem PCA- sem uso de qualquer tratamento matemático para reduzir a matriz de observações- terão apenas dados de amplitudes em determinadas frequências, escolhidas de forma aleatória (as mil primeiras calculadas por FFT), para reduzir a dimensionalidade dos dados. Estes testes seguirão a ordem:

1. Leitura dos dados oriundos da análise do FFT;
2. Separar os 1.000 primeiros dados em frequência (de 100 à 1.600 Hz, escolhidos de modo arbitrário);
3. Normalização dos dados;
4. Separação de dados apenas de um microfone;

Após estes quatro passos, seguem os passos de 4 a 13 das redes com dados após PCA, subseção 1.1.

A escolha das mil primeiras frequências é feita a partir de testes prévios de redes treinadas com outras faixas diferentes de frequência, escolhidas também aleatoriamente, por exemplo de 3.100 a 4.600 Hz. Os resultados com observações de outras faixas de frequências apresentaram respostas finais parecidas ou com limitações nas respostas. Esta análise, com as comparações de redes treinadas com observações de faixas diferentes de frequência, não será desenvolvida neste trabalho.

2.2) Redução da Dimensionalidade escolhendo apenas Dados em Regime Permanente

Neste teste, apenas dados em instantes próximos ao final do experimento, após 50 s, com aproximadamente 60 s serão usados com a ordem a seguir:

1. Leitura dos dados oriundos da análise do FFT;
2. Separar dados do primeiro microfone, m1, respectivos ao final da aquisição, aproximadamente 60 s;
3. Normalização dos dados;

Após estes três passos, seguiram-se os passos de 4 a 13 das redes com compressão, subsecção 1.1.

A Figura 3.4 possui um esquema geral da obtenção de dados e das investigações feitas. O ramo direto é a metodologia principal investigada e o ramo secundário indica os testes de metodologias alternativas sem a compressão de dados por PCA com possível potencial de identificar o sistema de vazamento.

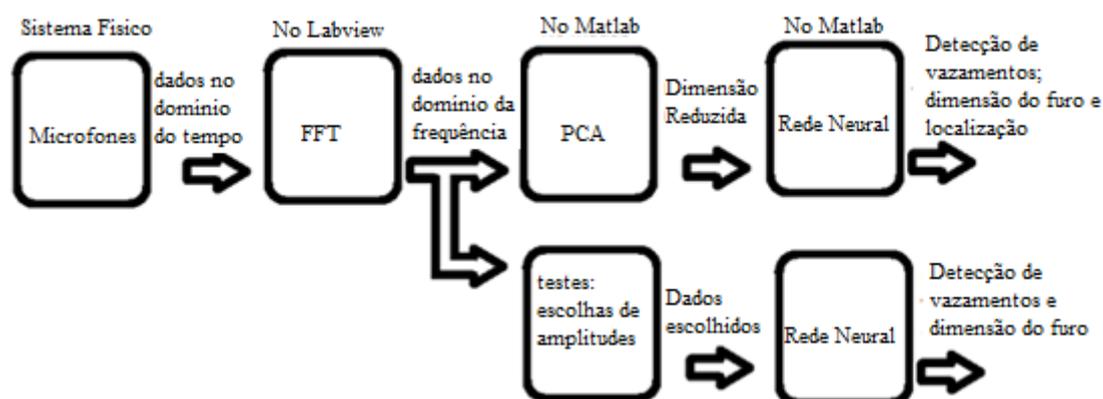


Figura 3.4 Esquema da aquisição e tratamento de dados. Ramo principal: metodologia investigada; ramo secundário, testes de metodologias alternativas

Nos testes sem a utilização de PCA, no ramo secundário, não houve investigação quanto a localização de vazamentos, posto que são testes apenas para avaliar o potencial das alternativas propostas e, de acordo com resultados, para serem aprofundadas em trabalhos futuros.

3.5 Planejamento dos Experimentos

Cada simulação terá duração de 60 s, no entanto, cada distúrbio só começará a partir de 10 s, ou seja, antes desse tempo inicial não há batidas, ou vazamentos em todas as configurações de experimento. Dados após a perturbação servirão como dados de entrada para o *Matlab*.

Um total de 15 experimentos é conduzido conforme Tabela 3.3.

Tabela 3.3 Experimentos Conduzidos

| Experimento | Conformação Experimental |
|-------------|--|
| 1 | Sem vazamento |
| 2 | Sem vazamento, com batidas manuais no botijão de armazenamento |
| 3 | Sem vazamento, com batidas manuais na tubulação |
| 4 | Vazamento no primeiro furo, F1 |
| 5 | Vazamento no segundo furo, F2 |
| 6 | Vazamento no terceiro furo, F3 |
| 7 | Vazamento no quarto furo, F4 |
| 8 | Vazamento no quinto furo, F5 |
| 9 | Vazamento no sexto furo, F6 |
| 10 | Vazamento no sétimo furo, F7 |
| 11 | Vazamentos simultâneos de F1 e F2 |
| 12 | Vazamentos simultâneos de F1 e F3 |
| 13 | Vazamentos simultâneos de F1 e F4 |
| 14 | Vazamentos simultâneos de F1 e F7 |
| 15 | Vazamentos simultâneos de F2 e F7 |

No experimento número 1, o sistema não enfrenta qualquer perturbação simulada.

Os experimentos 2 e 3 possuem batidas acrescentadas ao sistema para se acrescentar perturbações que não sejam vazamentos. Desta forma, a estratégia serve para diminuir a possibilidade de serem reconhecidas como vazamento, “falsos positivos”, na identificação do sistema.

Experimentos contendo um vazamento em um furo são realizados nos experimentos 3 a 10; Experimentos contendo dois vazamentos simultâneos serão realizados nos experimentos de 11 a 15.

4 RESULTADOS E DISCUSSÕES

Nesta seção, apresentam-se todos os resultados obtidos da investigação estudada de ocorrência de vazamentos na tubulação experimental.

4.1 Resultados de Aquisição

Os experimentos tiveram duração de 60 s. Durante a aquisição, os dados são armazenados em conjuntos de 20.000 em 20.000 pontos.

Os primeiros 20.000 pontos foram armazenados em 0,666 s; Mais 20.000 pontos foram armazenados no instante 1,332 s, com então 40.000 pontos de amplitude acumulados.

No domínio da frequência, a quantidade de pontos se observou ser sempre a metade do número de pontos do de tempo. Supostamente, então, se a aquisição parasse no instante 0,666 s, verificar-se-iam 10.000 pontos de amplitudes calculados no domínio da frequência e, no instante 1,332 s, mais 10.000 pontos, com 20.000 pontos acumulados.

Por exemplo, os dados armazenados contendo as amplitudes ao longo do tempo do experimento 4, de vazamento em F1, contém amplitudes até o tempo de 61,318 s. Ao todo, são 1.840.000 pontos de amplitudes no domínio do tempo observados para cada um dos cinco microfones que realizaram aquisições durante a simulação.

Enquanto que, nos dados armazenados para o mesmo experimento no domínio da frequência, há 920.000 dados calculados para cada microfone compreendidos nos valores de frequência de 0 Hz a 15.000 Hz, sendo que as amplitudes fora do intervalo de 100 Hz a 10.000 Hz estavam atenuadas pelo filtro e foram desconsideradas.

Considerando a velocidade de dados salvos como aproximadamente linear e constante, em regra de três simples, pode-se fazer uma estimativa de quais dados no domínio da frequência são respectivos a um determinado instante, ou intervalo de tempo.

No presente estudo, em vez de regra de três, ao se desejar saber quais as frequências lidas em determinado instante, repetiu-se a aquisição de dados até o respectivo instante de interesse e verificou-se qual o último conjunto de 10.000 dados

armazenados no domínio da frequência. Assim, esses pontos seriam respectivos àquele instante, ou intervalo de tempo.

Assim, apesar do cálculo da Transformada Rápida de Fourier não gerar informações da frequência juntamente com o tempo, pôde-se obter o tempo correspondente às observações em frequência.

4.2 Domínio do tempo

Os resultados no domínio do tempo dos experimentos com vazamento em, por exemplo, F1 e em F2 estão na Figura 4.1.

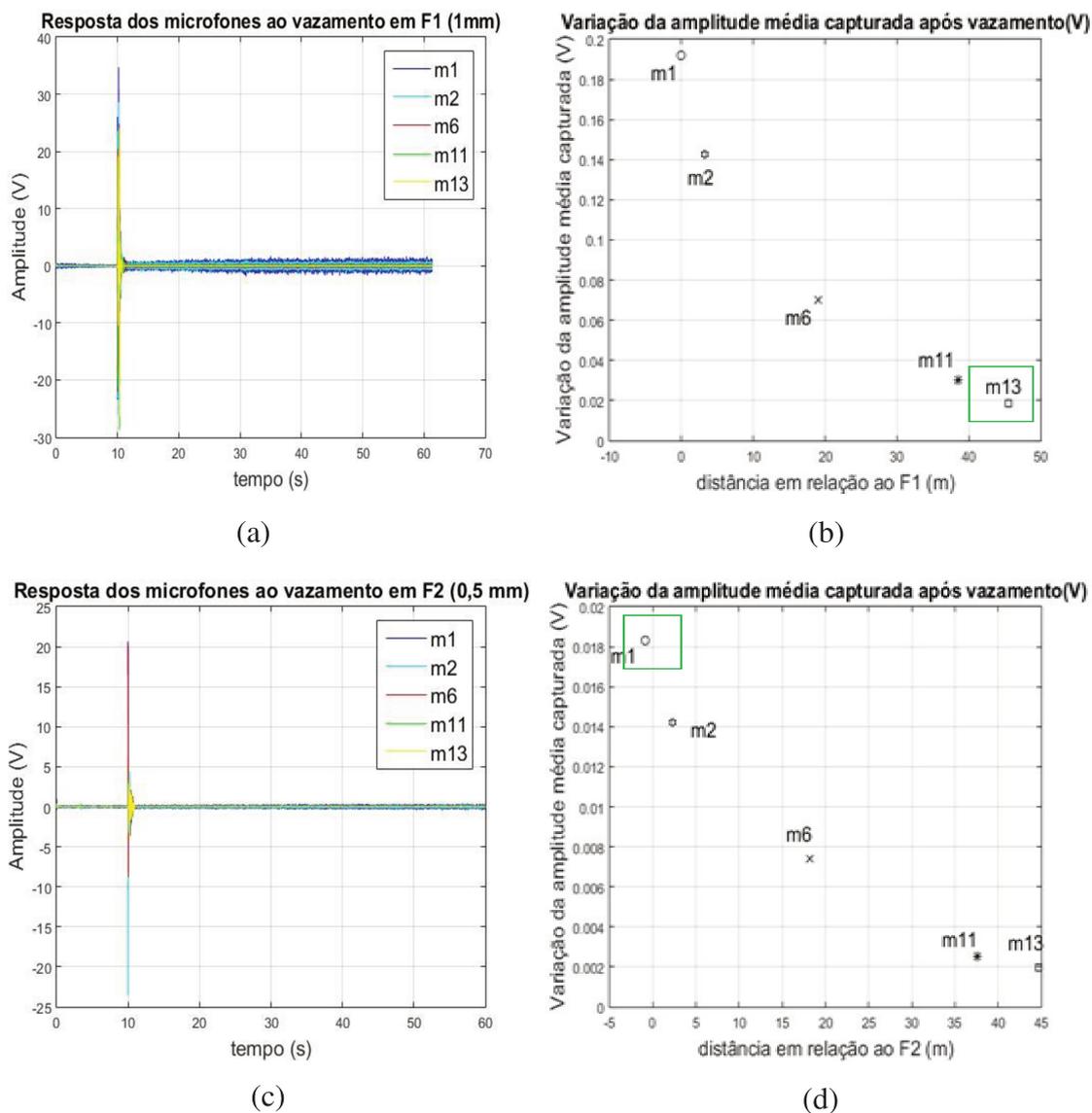


Figura 4.1 (a) Amplitude capturada ao longo do tempo em ocorrência de vazamento em F1; (b) Variação da amplitude nos microfones com a ocorrência do vazamento em F1; (c) Amplitude capturada ao longo do tempo em ocorrência de vazamento em F2; (d) Variação da amplitude nos microfones com a ocorrência do vazamento em F2

Percebe-se que, em (a) e (c), no momento em que o vazamento começa a ocorrer, em 10 s, os sinais dos microfones saem do intervalo entre -10 a 10 V. Também se observa que as amplitudes dos sinais aumentam com a ocorrência da perturbação. As figuras (b) e (d) da Figura 4.1 comprovam esse aumento mostrando uma variação positiva na amplitude acústica mesmo em microfones a mais de 30 m de distância da origem do vazamento.

O apêndice C apresenta o comportamento de outros experimentos no tempo. Neste, em conjunto com a Figura 4.1, percebe-se que após o trigésimo segundo do experimento, as amplitudes sustentam um comportamento (com valor mínimo e máximo) aproximadamente constantes em todas as configurações apresentadas de vazamento, sem mais variações no tempo.

Ao se comparar os gráficos das variações (b) e (d), a variação em todos os microfones é maior na ocorrência de vazamento em F1, no qual o diâmetro do furo é de 1,0 mm contra 0,5 mm de diâmetro no F2. Além disso, quanto mais próximo do microfone, maior esse aumento da amplitude. De fato, a variação na amplitude sonora no microfone m13 com a ocorrência de um vazamento de 1 mm de diâmetro e a uma distância de aproximadamente 45 m é aproximadamente a mesma variação ocorrida no microfone m1 à 0,9 m de um vazamento de 0,5 mm diâmetro. Estes dados estão evidenciados em retângulos verdes.

De qualquer modo, pode-se então, de modo rápido, identificar o microfone mais próximo de um vazamento, independentemente da dimensão do furo: aquele que possuir maior variação em sua amplitude sonora. Assim sendo, a comparação da variação das amplitudes dos microfones aponta diretamente para a localização aproximada do vazamento.

4.3 Domínio da Frequência

Ao se calcular o domínio da frequência, a partir do método FFT, os dados são decompostos em observações em frequências de 0 Hz a 16.000 Hz, o intervalo de capacidade do microfone utilizado. No entanto, amplitudes nas faixas fora do intervalo de 100 Hz e 10.000 Hz, intervalo em estudo, são atenuadas pelo filtro.

Para cada experimento, as amplitudes não filtradas e obtidas após a perturbação ocorrer foram guardadas e compuseram uma matriz de dados multidimensionais, pois cada amplitude calculada dentro do intervalo de 100 Hz a 10.000 possuía observações ao longo do tempo.

O sistema experimental estava exposto a perturbações externas e não controladas, barulhos no ambiente como de passos, do ar-condicionado em funcionamento etc., são exemplos. Assim, as respostas obtidas no domínio da frequência não eram estáticas, mas possuíam variações de um instante a outro, mesmo que pequenas.

A Figura 4.2 está dividida em representações das frequências capturadas em intervalos de tempo estimados diferentes no experimento de vazamento em F1.

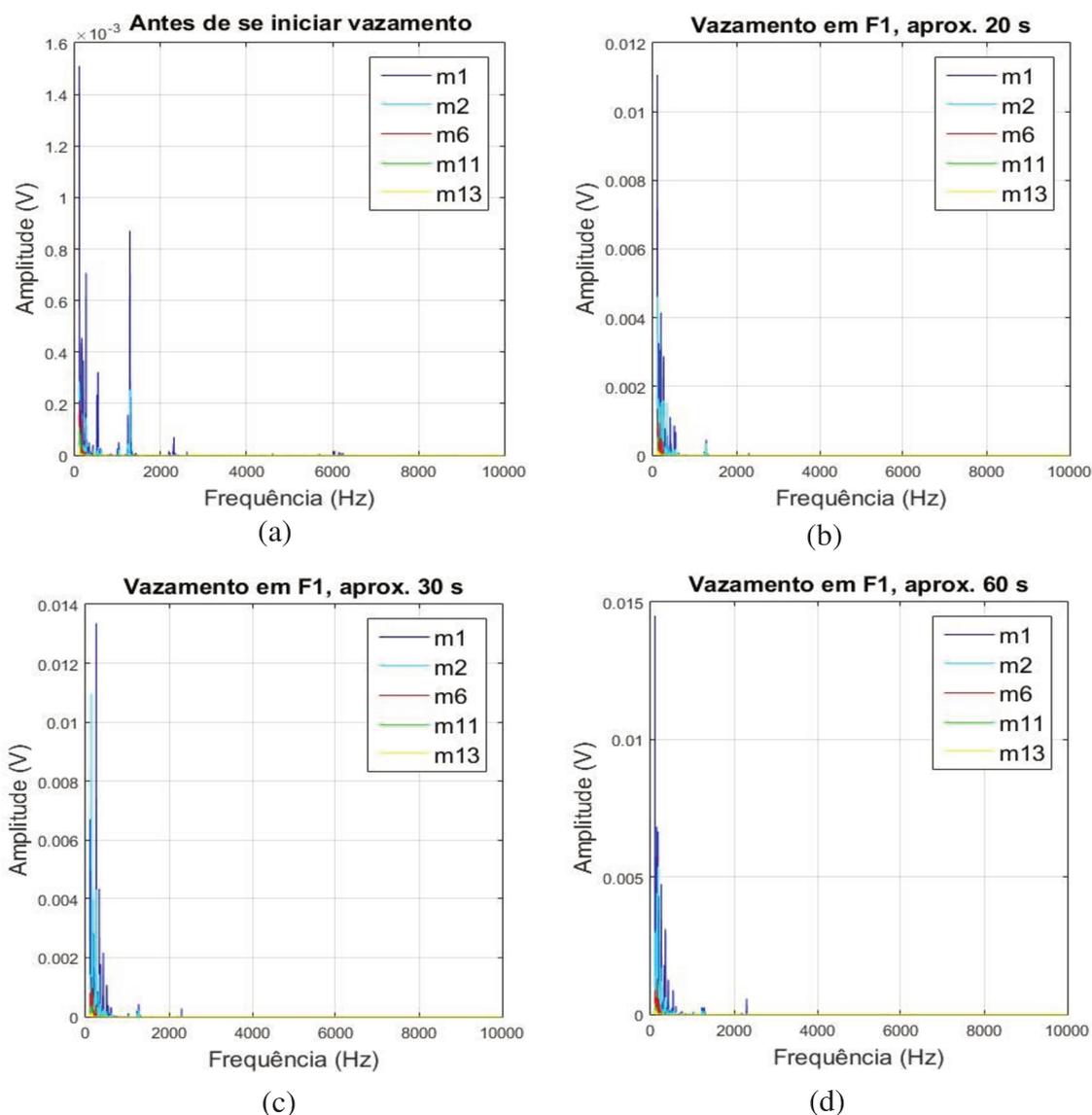


Figura 4.2 Vazamento em F1 em tempos aproximados diferentes. (a) antes de simular vazamento; (b) em aproximadamente 20 s; (c) em aproximadamente em 30 s; (d) em aproximadamente em 60 s

Pela Figura 4.2, novamente, verifica-se que de 20 à 60 s de experimento não há muitas variações perceptíveis. No presente estudo, assume-se que em 60 s o sistema já esteja com os sinais acústicos em regime permanente e que as possíveis modificações sejam atribuídas às perturbações não controladas do sistema. Assim, os dados capturados neste estudo representam tanto o regime transiente, quanto o regime

permanente do sistema com as perturbações. O Apêndice D apresenta frequências no tempo de outras configurações de vazamentos com maior diâmetro.

A matriz de dados de frequência na qual foi realizada a Análise de Componentes Principais possuía todas as amplitudes de 100 a 10.000 Hz calculadas por FFT a partir de 10 s (quando se iniciavam as perturbações no sistema, se houvesse, batidas ou vazamentos) de todas as configurações de vazamento, de todos os microfones, até 60 s. Assim, uma matriz multidimensional, em que amplitudes em determinada frequência é uma variável com observações. A dimensão da matriz resultante da junção dos dados referentes a todos os experimentos, a partir de 10 s, é de 5.475×6.599 (5.475 observações de amplitudes em cada uma das 6.599 frequências diferentes dentro do intervalo de 100 a 10.000 Hz).

4.4 Análise de Componentes Principais

Os dados, antes de se realizar a compressão, foram padronizados para que as observações de amplitudes de cada frequência obtida tivessem média 0 e desvios padrões de 1.

A análise de Componentes Principais tem o intuito de diminuir a dimensionalidade dos dados de frequência conservando ao máximo as informações principais que os dados representam. O teste do Percentual Cumulativo de Variância (PCV) está representado pela Figura 4.3.

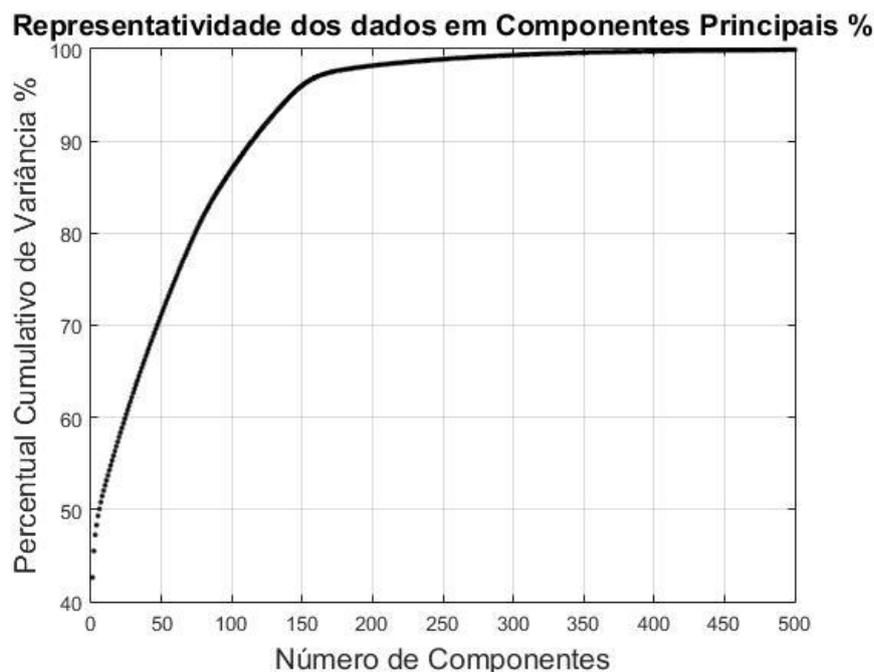


Figura 4.3 Teste Cumulativo de Variância dos dados de frequência em termos de componentes principais

Como resultado da Análise de Componentes Principais, a partir de teste percentual cumulativo de variância, utilizando-se apenas 3 componentes principais, ou seja, comprimindo os dados a apenas 3 dimensões (5.473x3, 5.473 observações em 3 dimensões), obtêm-se 47,26 % da variância original dos dados. Em 200 componentes, 98,19 % e, em 500, 99,91 %.

Apesar de 3 dimensões, para este conjunto de dados, representarem apenas 47 % da variância original, estas são consideradas o suficiente para uma análise prévia dos dados. O gráfico, dos dados em termos das 3 componentes principais (C.P.), representa, em partes, a distribuição dos dados após a análise de componentes principais e a visualização de informações antes não explícitas.

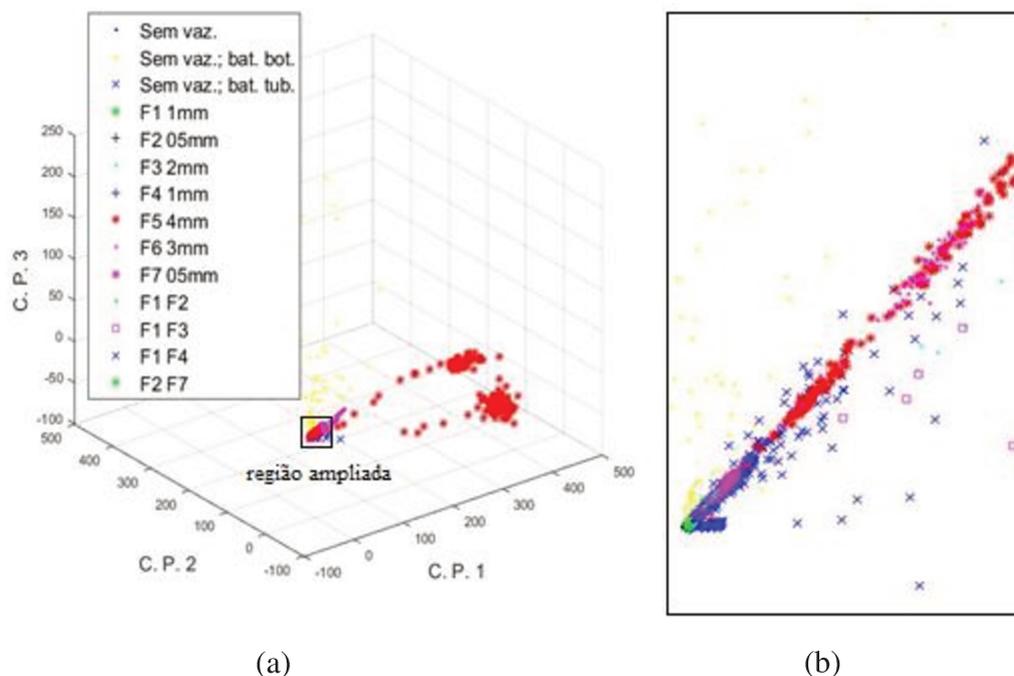


Figura 4.4 (a) Todos os dados experimentais em 3 dimensões; (b) Região ampliada

A partir da Figura 4.4, sugere-se que os dados de um mesmo experimento tendem a se agrupar, porém, um mesmo experimento apresenta mais de um agrupamento. A exemplo, visto com mais nitidez na Figura 4.4 (a), vazamento em F5 (de diâmetro de 4 mm), possui ao menos dois agrupamentos distintos, em vermelho. Percebe-se também que F5 possui um padrão distinto das demais configurações.

Para investigar o significado de uma mesma configuração experimental possuir mais de um agrupamento, a Figura 4.5 representa situações de um vazamento apenas, distinguindo os microfones.

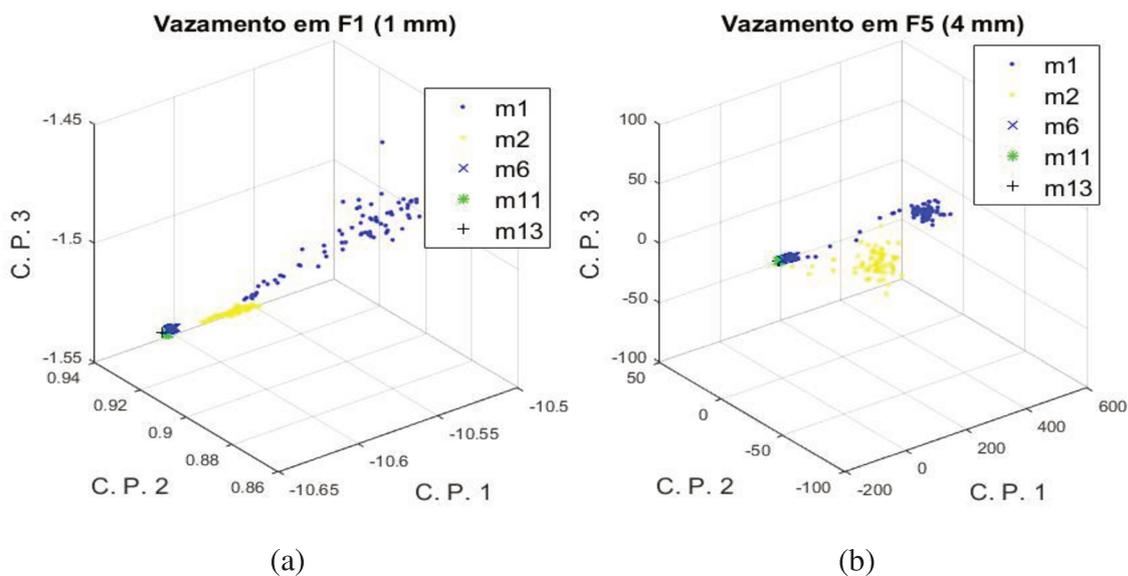


Figura 4.5 Dados em 3 dimensões de dados de apenas um experimento: (a) De vazamento em F1; (b) de vazamento em F5

De acordo com a Figura 4.5, pode-se associar os diferentes agrupamentos de um mesmo experimento a respostas dadas por diferentes microfones, que se encontram em posições diferentes ao longo da tubulação de simulação. O padrão distinto de F5, percebido na Figura 4.4 e na Figura 4.5 (b), também pode ser mais especificado como pertencendo às respostas dos microfones 1 e 2, mais próximos do vazamento.

Com o gráfico em 3 componentes principais, pode-se observar que cada configuração de experimento tende a se distinguir no gráfico e, além disso, cada microfone também possui seus padrões. Porém, a análise em apenas 3 componentes principais não é suficiente para uma boa representatividade dos dados, visto que 200 componentes principais seriam uma compressão bem mais significativa compreendendo mais de 98% da variância original.

4.5 Rede Neural de Kohonen

É considerado que 98% de representatividade seja suficiente para o presente estudo. Assim, 200 componentes principais, com 98,19 % de variância explicada, são utilizadas. Para esta análise com 200 dimensões, utilizou-se a rede neural não supervisionada de Kohonen.

Antes do início do treinamento, foi também realizada uma normalização dos dados com a função *som_normalize*, assim, os dados, em termos de componentes principais, foram novamente padronizados. Com esta função, uma alternativa de método de normalização é a de equalização de histograma, geralmente utilizada quando há diferenças muito grandes nos dados, comumente usada em tratamento de imagens.

Para o treino de redes auto-organizáveis no *Matlab*, a função *som_make* é capaz de inicializar e de treinar dados sem qualquer especificação direta para a configuração da rede. Esta função determina a quantidade de neurônios a serem utilizados de acordo com o tamanho da matriz dos dados de treino por uma equação empírica, assim como o número de iterações para cada fase do treinamento; enquanto as funções *default* são utilizadas para as funções de vizinhança e de taxa de aprendizado (gaussiana e inversa). E, para os parâmetros que apareçam nas funções, valores padrões também são assumidos etc [34].

Assim, para todos os treinamentos realizados, primeiramente a função *som_make* era utilizada como ponto de partida. Algumas modificações eram testadas no intuito de se melhorar o treinamento. Dentre estas modificações, eram testados diferentes números de neurônios, métodos de normalizações, funções de vizinhança etc. Os valores dos erros topográficos e de quantização, além da organização final da Matriz U gerada, eram comparados para se encontrar o melhor conjunto de configurações para a rede em questão.

4.5.1 Redes Treinadas com Dados Comprimidos por PCA

Nesta seção estão expostos os resultados das redes não supervisionadas com dados de entrada em termos de suas componentes principais.

Na Tabela 4.1, estão os métodos de normalização, os erros de quantização (E. Q) e topográfico (E. T.) e o objetivo de cada rede. A inicialização dos pesos de todos os treinamentos foi linear.

Tabela 4.1 Redes realizadas

| Rede | Microfone | Normalização | Tamanho da Rede | E. Q. | E. T. | Objetivo da Rede |
|------|-----------|---------------------------|-----------------|-------|-------|---|
| 1 | Todos | Equalização de histograma | 40x42 | 2,026 | 0,010 | Localizar microfones em distâncias diferentes |
| 2 | M1 | Eq. de histograma | 17x8 | 3,353 | 0,001 | Determinar dimensão de vazamento |
| 3 | M2 | Eq. de histograma | 17x8 | 2,931 | 0,003 | Determinar dimensão de vazamento |
| 4 | M6 | Eq. de histograma | 22x10 | 1,693 | 0.023 | Determinar dimensão de vazamento |
| 5 | M11 | Eq. de histograma | 25x10 | 1,886 | 0,012 | Determinar dimensão de vazamento |
| 6 | M13 | Eq. de histograma | 28x6 | 1,241 | 0.051 | Determinar dimensão de vazamento |

A Tabela 4.2 mostra os parâmetros de treinamento das redes. Estes foram determinados por tentativa e erro de acordo com a diminuição dos erros topográfico e de quantização, levando também em consideração a organização final da matriz U gerada.

Tabela 4.2 Parâmetros de Treinamento das Redes

| Rede | Função $\alpha(t)$ | Função de vizinhança | Algoritmo | Vizinhança dos neurônios | Topologia da Rede |
|------|--------------------|----------------------|------------|--------------------------|-------------------|
| 1 | - | <i>Cutgaussian</i> | Batelada | hexagonal | Plana |
| 2 | <i>Power</i> | <i>Gaussian</i> | Sequencial | hexagonal | Plana |
| 3 | - | <i>Gaussian</i> | Batelada | hexagonal | Plana |
| 4 | - | <i>Cutgauss</i> | Batelada | hexagonal | Plana |
| 5 | - | <i>Gaussian</i> | Batelada | hexagonal | Plana |
| 6 | - | <i>Bubble</i> | Batelada | hexagonal | Plana |

1) Construção da Rede 1: Todos os Dados

Tendo como base a análise preliminar em apenas 3 dimensões, verificou-se que microfones a distâncias diferentes da origem da perturbação possuem padrões diferentes. Assim, uma primeira rede é realizada para determinar a distância aproximada da origem do vazamento ao microfone, ou seja, distinguir os dados por microfones.

A primeira rede construída possui em seu treinamento todos os dados oriundos da análise de Componentes Principais. Os dados com observações se encontravam em uma matriz de dimensões de 5.475x200 (dados comprimidos a 200 dimensões). Foram reagrupados em 2 matrizes, uma para treinamento da rede de dimensões de 3.750x200 e uma matriz com dados para teste de 1.725x200.

Na Figura 4.6, os neurônios- vetores pesos resultantes do treinamento pelo algoritmo em batelada, equação 2.9- estão representados como pontos azuis.

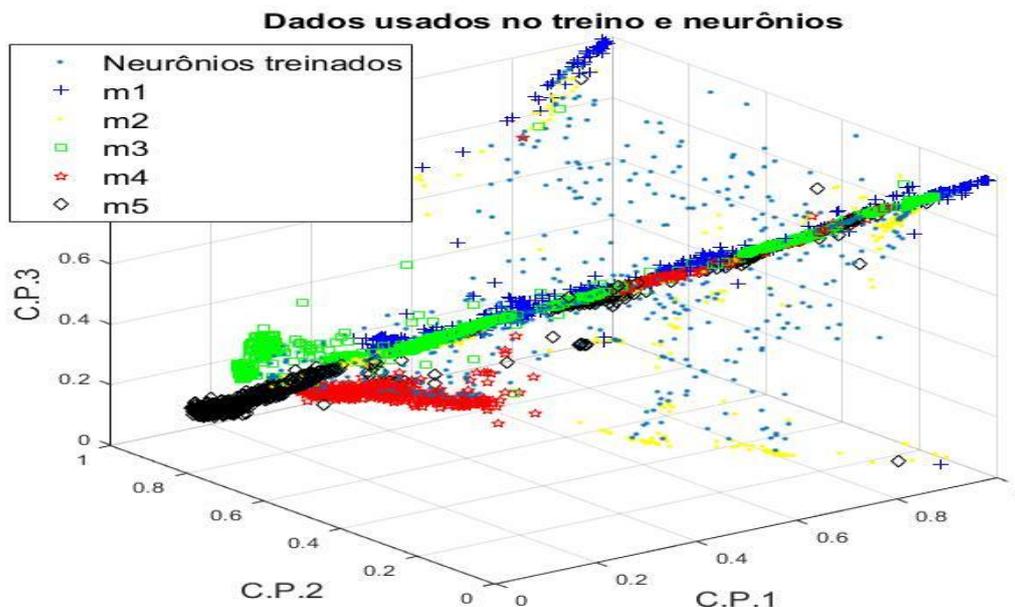


Figura 4.6 Neurônios treinados e os dados de treinamento padronizados, em termos de suas componentes principais (C. P.)

O gráfico da Figura 4.6 é próximo àquele apresentado na Figura 4.4 (a), porém após observações serem normalizadas pelo método de equalização de histograma e acrescentando os neurônios treinados. Além disso, os dados estão evidenciados em termos de microfones e não de configuração do sistema.

Os neurônios treinados possuem o mesmo número de dimensões que os dados utilizados em seu treinamento. Neste caso, assim, os neurônios possuem 200 dimensões. Então, por utilizar apenas três componentes principais, os neurônios também possuem 197 dimensões não representadas na Figura 4.6.

A matriz U deste treinamento está representada na Figura 4.7 A construção da matriz U leva em consideração todas as dimensões do neurônio.

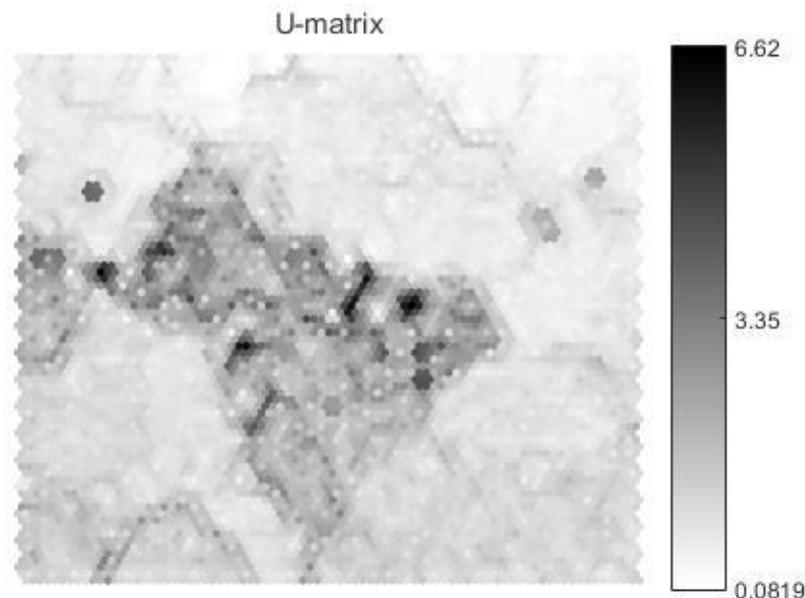


Figura 4.7 Representação dos Neurônios em matriz U

Na Figura 4.7, a escala ao lado indica a proximidade de um hexágono ao outro. O Apêndice A possui um exemplo de construção de matriz U e de sua leitura. Nesta matriz U, dificilmente se distinguem grupos de neurônios relacionados a microfones diferentes. Para tal, verificaram-se quais neurônios são ativados pelos dados de cada microfone.

Na Figura 4.8 (a) evidenciam-se os hexágonos, representando neurônios, ativados pelos dados de treinamento de microfones diferentes e, em (b), são agrupados visualmente áreas correspondentes a um mesmo microfone (mesma coloração). Em (c), são evidenciados os hexágonos, correspondente a neurônios, ativados pelos dados de teste e, em (d), são agrupados visualmente áreas ativadas por mesmo microfone.

Percebe-se que os dados de teste ativam, em sua maioria, neurônios em mesma área que os dados de treino, estabelecendo um bom teste para a rede treinada. Assim, a partir da área ativada pelos dados, pode-se estimar a distância do microfone à origem da perturbação.

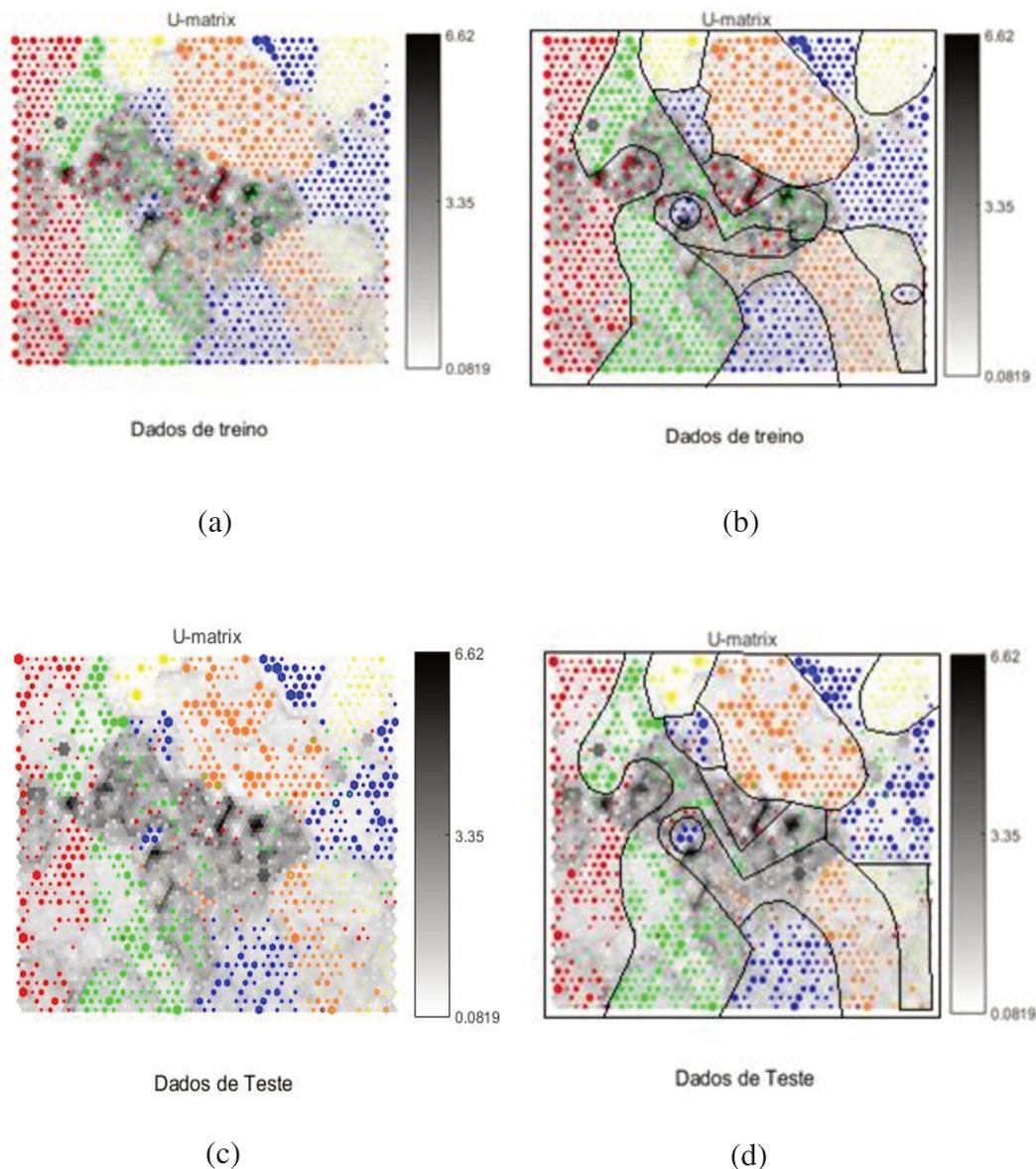


Figura 4.8 Matriz U evidenciando ativação dos neurônios a microfones em posições distintas. (a) Matriz U evidenciando neurônios ativados por dados de treino; agrupados visualmente em (b). (c) Matriz U evidenciando neurônios ativados por dados de teste; agrupados visualmente em (d). Vermelho: m1; verde: m2; laranja: m6; azul: m11 e amarelo: m13.

Nesta mesma rede, é provável também se conseguir definir qual a perturbação, vazamento ou batidas, em ocorrência. A exemplo, os tipos de distúrbios (batidas, ou 1 vazamento apenas) em ocorrência, capturado pelo 2º microfone, m2, microfone nas proximidades dos vazamentos, e pelo 13º, m13, o microfone mais distante em estudo, estão evidenciados na Figura 4.9 e na Figura 4.10.

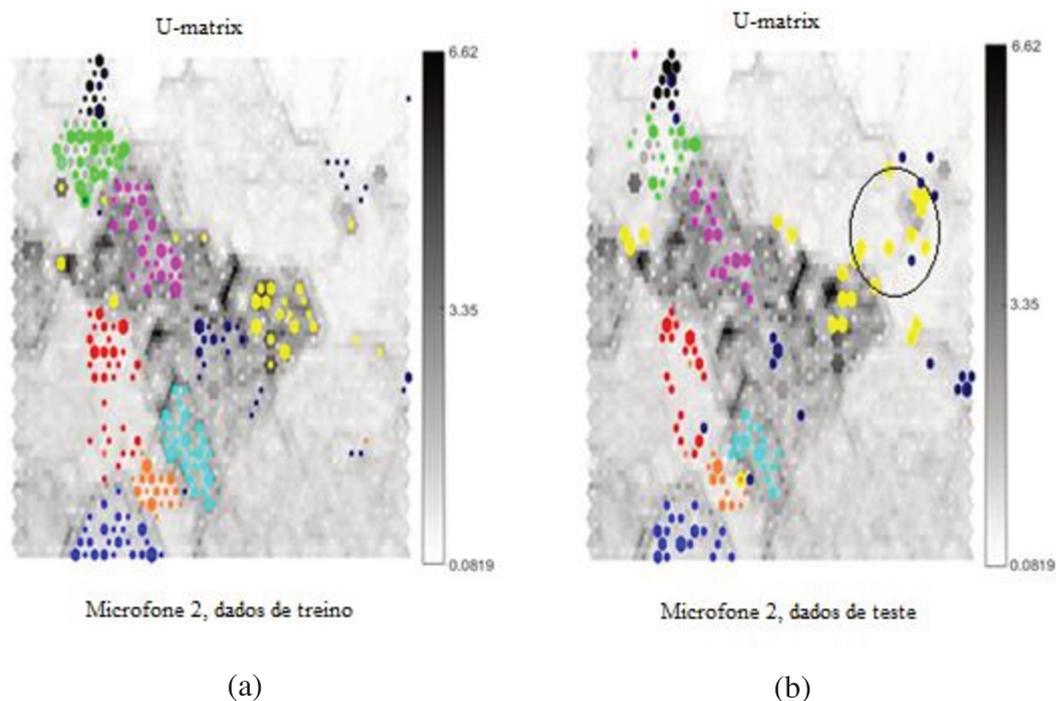


Figura 4.9 Matriz U evidenciado ativação dos neurônios aos dados de m2 apenas. (a) Dados de treino; (b). Dados de teste. Situações sem vazamentos- preto: sem anormalidades; rosa escuro: batidas no botijão; azul escuro: batidas na tubulação. Com vazamentos- vermelho: em F1 (1 mm); verde: em F2 (0,5 mm); azul: em F3 (2 mm); laranja: em F4 (1 mm); amarelo: em F5 (4 mm); ciano: em F6 (3 mm); e cinza: em F7 (0,5 mm)

Na Figura 4.9 (b), verifica-se que a distribuição de hexágonos em amarelo, representando neurônios ativados por dados de teste de vazamento em F5 (furo de maior diâmetro, 4 mm), são os que mais se distinguem dos dados de treinamento. A área evidenciada no teste, circulada, não aparece em amarelo pelos dados de treinamento. Uma possível causa para isto seria pelo fato de F5 ser o maior vazamento. Este tem, por hipótese, um regime transiente mais longo, o que pode ser a causa da diferença nos padrões dos dados separados para treino e para teste.

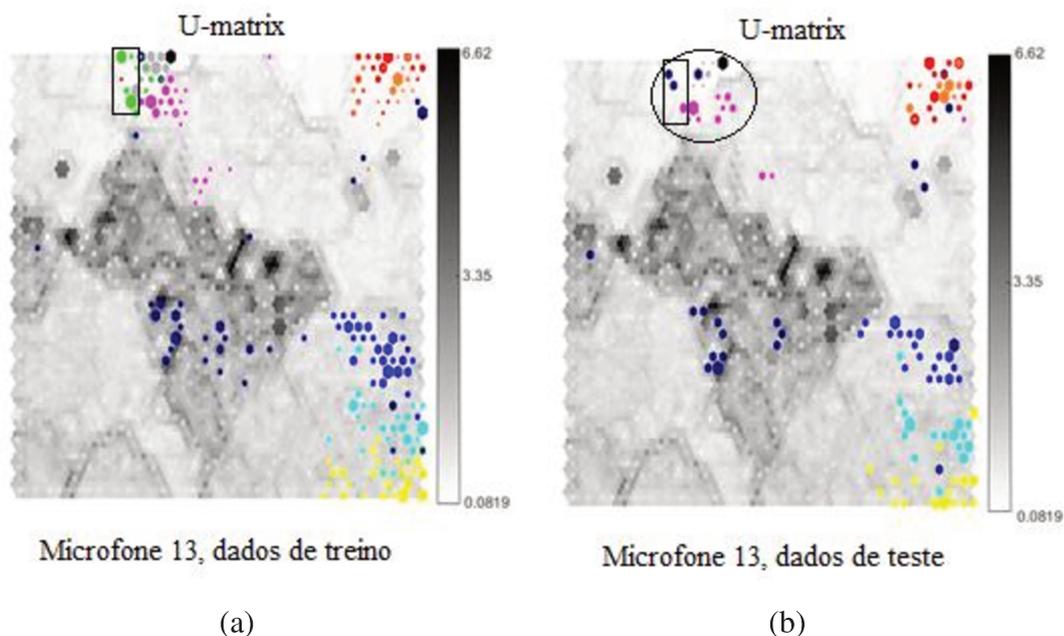


Figura 4.10 Matriz U evidenciado ativação dos neurônios aos dados de m13 apenas. (a) Dados de treino; (b). Dados de teste. Situações sem vazamentos- preto: sem anormalidades; rosa escuro: batidas no botijão; azul escuro: batidas na tubulação. Com vazamentos- vermelho: em F1 (1 mm); verde: em F2 (0,5 mm); azul: em F3 (2 mm); laranja: em F4 (1 mm); amarelo: em F5 (4 mm); ciano: em F6 (3 mm); e cinza: em F7 (0,5 mm)

Quanto mais perto do vazamento, com mais detalhes se mostram as respostas dos microfones: a exemplo, as mudanças nos agrupamentos em amarelo não foram acentuadas em m13, Figura 4.10, como foi em m2, Figura 4.9.

Quanto maior o número de dados de treinamento, menos precisa pode ser a resposta da rede, ou seja, mais suscetível a erros. Percebe-se na Figura 4.10 (b) que há dificuldades no teste de determinar se ocorre vazamento de 0,5 mm ou se não há vazamentos, circulado ao topo. Mais especificamente, nos retângulos, neurônios ativado por dados de treino de vazamento de 0,5 mm de diâmetro (verde e cinza), são ativado por dados de teste de sem vazamento com batidas no botijão e na tubulação (rosa e azul escuros).

Assim, para uma análise mais específica, optou-se por primeiramente utilizar a Rede 1, com os dados de todos os microfones, apenas para definir a distância aproximada, e, posteriormente, outras redes, identificam o distúrbio em questão.

2) Construção da Rede 2: Dados do Microfone 1

Para treino da Rede 2, apenas dados do microfone 1, m1, foram considerados. O microfone 1, no sentido de fluxo do ar comprimido, está localizado antes dos furos de vazamentos, a 0,1 m de F1 e a 5,5 m de F7, o vazamento mais distante deste microfone. O objetivo da rede é distinguir as perturbações ocorridas ao sistema.

Os dados apenas do microfone 1, que é submatriz da matriz de dados usados para treino e teste da Rede 1, formavam uma matriz de dimensões de 1.095x200 (1095 observações em 200 componentes principais). Destes dados, 750 observações foram separadas para treino e 345, para teste.

O gráfico representando apenas as 3 primeiras dimensões, em componentes principais, dos dados de treinamento do microfone 1 e dos neurônios treinados- vetores pesos resultantes do treinamento pelo algoritmo sequencial, equação 2.6- encontra-se na Figura 4.11.

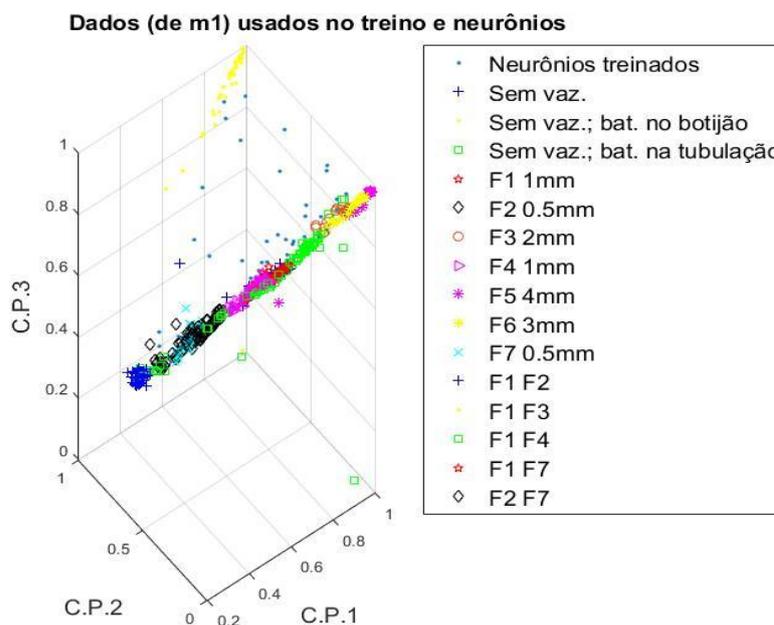


Figura 4.11 Neurônios treinados e os dados de treinamento de m1 padronizados, em termos de suas componentes principais (C. P.)

A representação dos neurônios treinados em matriz U está na Figura 4.12.

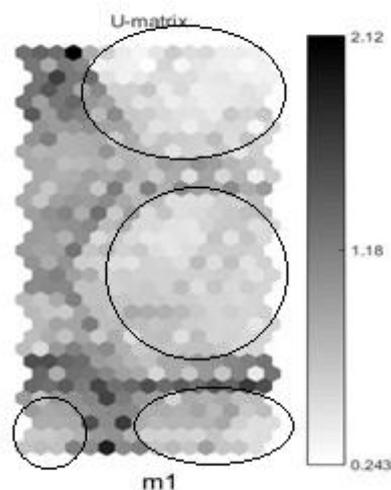


Figura 4.12 Representação dos neurônios em matriz U com dados apenas de m1

Áreas claras representam neurônios próximos, com características afins; partes escuras, características distintas. As regiões circuladas são regiões de neurônios próximos entre si, ou seja, são provavelmente ativados por mesmos padrões ou similares.

Novamente, para verificar quais neurônios estão associados aos diferentes padrões estudados do sistema, foi gerada a Figura 4.13. De (a) a (d) são informados os neurônios ativos em cada configuração pelos dados de treino enquanto em (e) e (f) são os neurônios ativados em cada configuração pelos dados de teste.

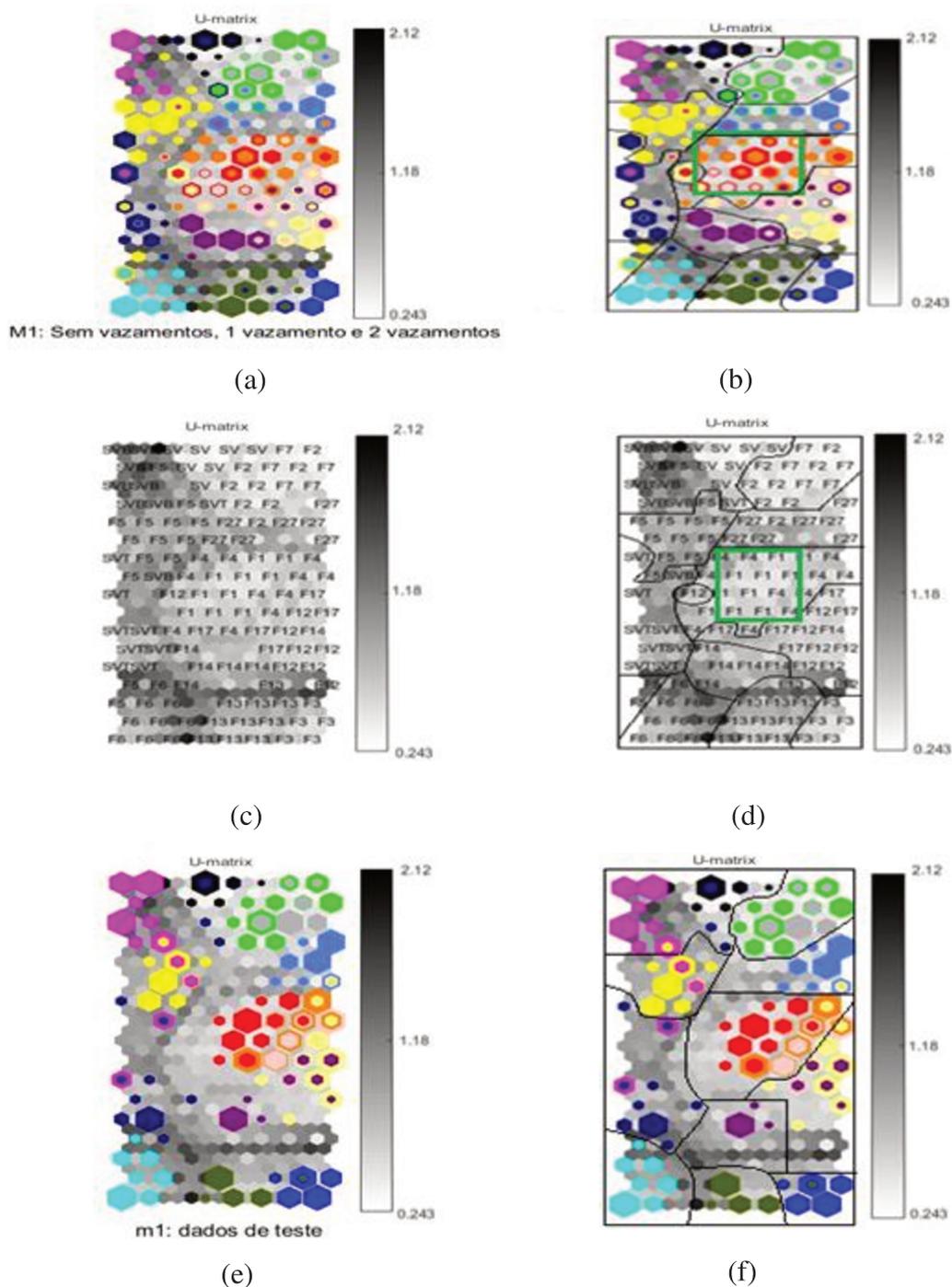


Figura 4.13 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27)

As representações por cores na Figura 4.13 possuem algumas características como: quanto maior o hexágono com certa coloração, mais vezes o respectivo neurônio foi ativado por dados da configuração do sistema representada por esta cor; se mais de uma cor está representada no hexágono, mais de uma configuração ativa o mesmo neurônio.

Nos agrupamentos feitos nas figuras à direita, o neurônio ativado representado com mais de uma cor era atribuído ao grupo da cor que ocupava maior área no hexágono. Enquanto que, nas figuras em que a matriz U apresentava legendas nos hexágonos ((c) e (d)), estas legendas representavam apenas a configuração que mais ativou cada neurônio. Assim, o agrupamento era facilitado, porém não gerava nenhuma informação se mais de uma configuração ativava o mesmo neurônio.

As figuras (a) e (c) são representações diferentes do mesmo resultado: ativação dos neurônios pelos dados de treinamento e (b) e (d) são os agrupamentos feitos visualmente de cada configuração estudada.

O resultado da matriz U nas figuras (b) e (d) da Figura 4.13 mostram que, para este treinamento, cada configuração (sem vazamentos, vazamentos de 1 furo e vazamentos de 2 furos) possui uma área própria de neurônios ativados.

Por exemplo, a região que contém o retângulo verde compreende uma região ativada com vazamentos de 1 mm de diâmetro, figuras (b) e (d). Na matriz U com legendas, figura (d), é fácil de perceber que vazamentos em F1 e em F4 são os que mais ativam esta área. No entanto, na representação com as cores, observa-se que há ativação desses neurônios, em menor escala, também por vazamentos simultâneos de F1 e F2 (amarelo claro, F12) e de F1 e F7 (rosa claro, F17) simultâneos, ou seja, vazamentos de 1 mm e de 0,5 mm ao mesmo tempo. Percebe-se que F1 e F7 simultâneos são os que mais vezes ativam estes neurônios, pela maior presença de hexágonos em rosa claro do que de amarelo claro nesta área, esta proximidade significa similaridades nos padrões.

Porém, vazamentos simultâneos de 1 mm e de 0,5 mm (F12 e F17) também possuem uma área própria na qual é a configuração principal de ativação. Esta é a área abaixo do retângulo verde.

A ativação dos neurônios pelos dados de teste (e) apresentou uma distribuição de cores próxima àquela aos dados de treino. Os agrupamentos, em (f), foram bastante parecidos e, em geral, pode-se observar que as mesmas configurações ativaram os mesmos neurônios.

Deste modo, pode-se observar qualitativamente que ocorre a ativação de uma área de neurônios, um agrupamento, que corresponde a um tipo de configuração de vazamento, de batidas, ou situação sem perturbação. Assim, identificando o sistema.

3) Construção da Rede 3: Dados do Microfone 2

Para treino da Rede 3, apenas dados do microfone 2, m2, foram considerados. Este microfone se encontra fisicamente entre os furos de vazamentos simulados, entre F4 e F5. O objetivo desta rede é a de distinguir as perturbações ocorridas ao sistema.

Apenas os dados do microfone 2 foram utilizados. Esta matriz, assim como a do microfone 1, possuía dimensões de 1.095 x 200 (1095 observações em 200 componentes principais). Destes dados, 750 observações foram separadas para treino e 345, para teste.

A representações em 3 dimensões dos dados de treino do microfone 2 e dos neurônios após treinamento- vetores pesos resultantes do treinamento pelo algoritmo em batelada, equação 2.9- estão na Figura 4.14.

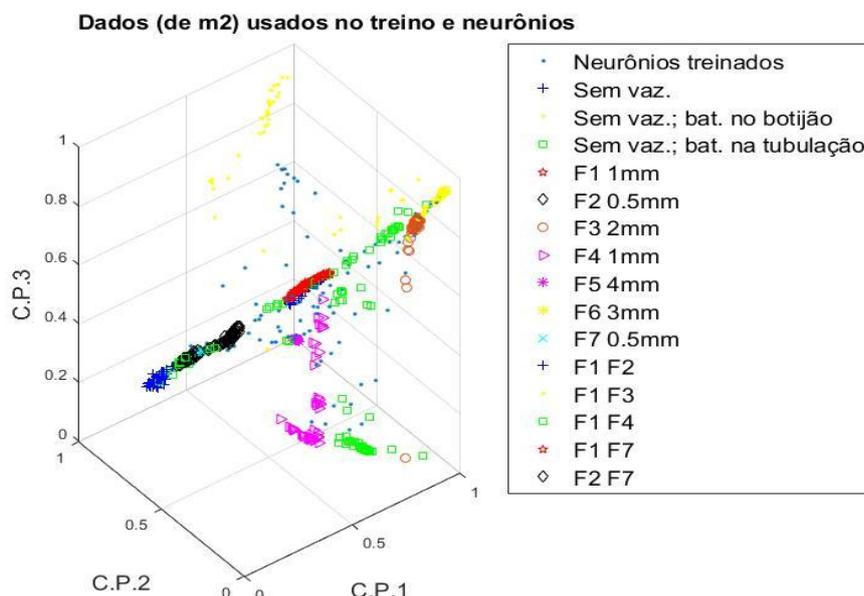


Figura 4.14 Neurônios treinados e os dados de treinamento padronizados de m2, em termos de suas componentes principais (C. P.)

A Figura 4.15 apresenta a representação em matriz U deste treinamento.

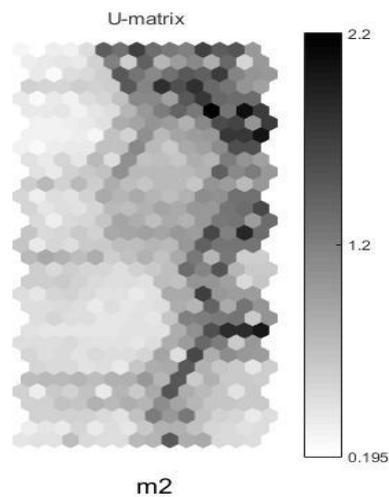


Figura 4.15 Representação dos neurônios em matriz U com dados apenas de m2

Os neurônios ativados por dados de treino e teste estão evidenciados na Figura 4.16. De (a) a (d) são os calculados os neurônios ativos em cada configuração pelos dados de treino, enquanto em (e) e (f) são os neurônios ativados em cada configuração pelos dados de teste.

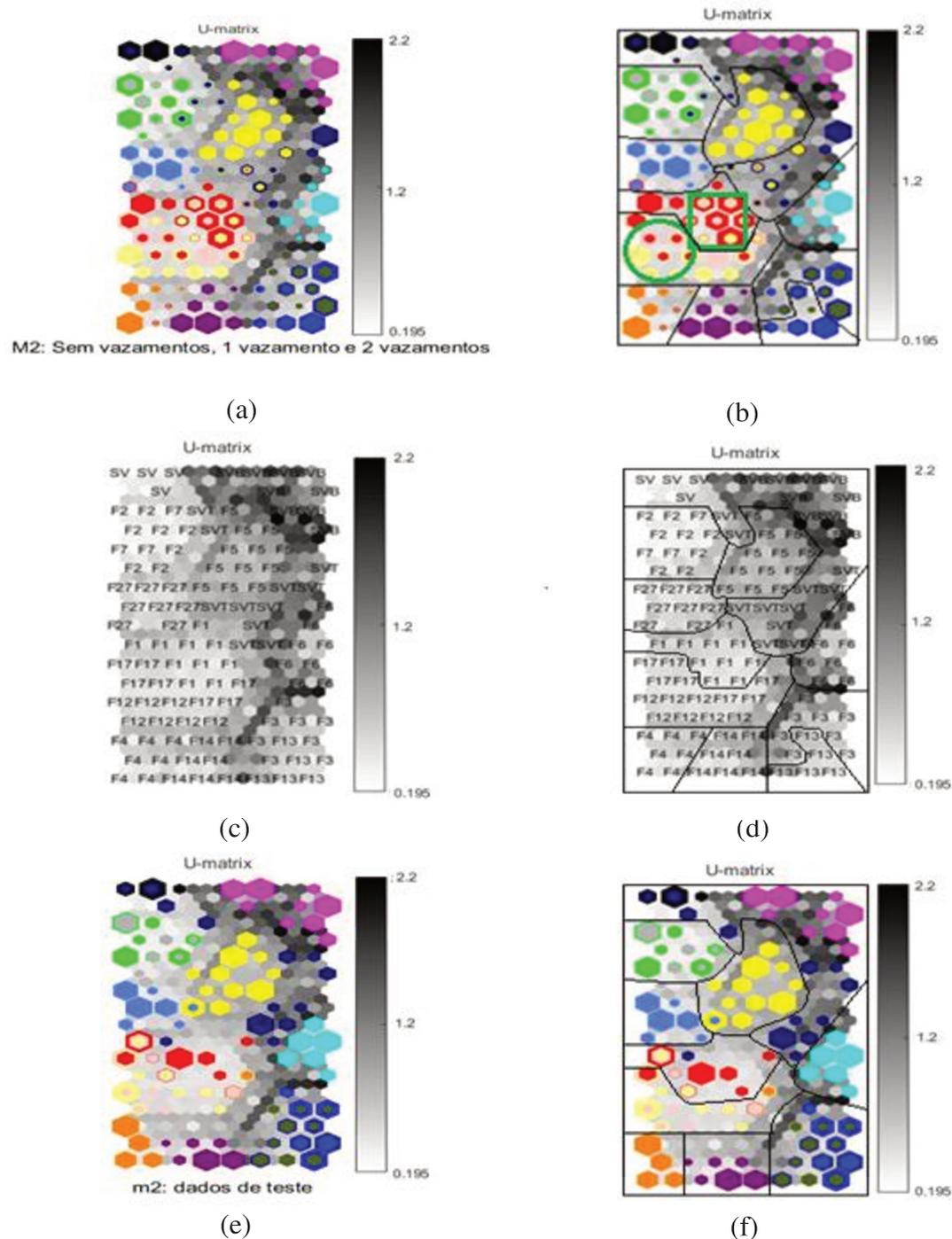


Figura 4.16 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27)

A Figura 4.16 foi construída da mesma maneira que a figura equivalente para o microfone 1, a Figura 4.13. Em seguida, para os microfones m6, m11 e m13 também há uma figura neste formato.

Observando-se os mapas (b) e (d) da Figura 4.16, percebe-se que para este treinamento com dados do microfone 2 também há a distribuição de áreas de neurônios que são ativados pela mesma configuração.

Diferentemente do resultado de m1, F1 (em vermelho) e F4 (em laranja), ambos vazamentos de 1 mm, foram separados em 2 grupos distintos, figura (b) e (d). E, de modo semelhante, o grupo atribuído à F1, em evidência por retângulo verde em (b), sofre também ativações de vazamentos simultâneos de 1 mm e 0,5 mm, observado pela presença de hexágonos rosa claro (vazamento em F1 e F7) e amarelo claro (F1 e F2) no interior de hexágonos maiores vermelhos (vazamento em F1).

Por outro lado, o agrupamento de vazamentos de 1 mm e de 0,5 mm, evidenciado pelo círculo em verde em seu interior, também possui ativações de dados de F1, observado pela presença de hexágonos vermelhos no interior de hexágonos maiores em rosa claro e em amarelo claro neste grupo, figura (b).

No treinamento, foram distinguidos em dois grupos os vazamentos em F3 e vazamentos simultâneos para F1 e F3, figura (b). Porém, no teste, a distinção não foi feita, agrupando hexágonos em azul (F3) e em verde-escuro (F1 e F3 simultâneos) num mesmo grupo, figura (f). Uma possível causa para uma maior dificuldade em dividir esses grupos pode ser causada pelo fato do microfone 2 se encontrar mais próximo de F3 do que de F1. Nesta hipótese, a perturbação causada por um vazamento de 2 mm de diâmetro próximo ao microfone (a 1,7 m de distância) pode esconder a perturbação causada por um vazamento de 1 mm um pouco mais distante (a 3,2 m de distância) do microfone. Enquanto que o microfone 1, que conseguiu distinguir os dois vazamentos tanto no teste quanto no treino, se encontra mais próximo de F1 (a 0,1 m de distância) do que de F3 (a 1,5 m de distância).

Comparando as figuras (e) e (f), com os agrupamentos, com (a) e (b), percebem-se muitas similaridades na organização das cores e dos agrupamentos feitos, o que configura um bom treinamento. Desta forma, configurações de vazamentos, ou perturbações em geral, podem ser associadas a áreas neurônios.

Assim, de acordo com a área ativada, também, se identifica o que ocorre no sistema.

4) Construção da Rede 4: Dados do Microfone 6

Para treino da Rede 4, desta vez, apenas dados do microfone 6, m6, foram considerados. Este microfone se encontra fisicamente após os furos de vazamentos simulados. A Rede 4 tem o objetivo de distinguir as perturbações ocorridas ao sistema. O microfone 6 está a 19,2 m de distância da entrada de ar comprimido na tubulação, a 19,0 m de distância de F1, o vazamento simulado mais longe deste microfone, e a 13,6 m de F7, o vazamento mais próximo simulado.

As dimensões da matriz contendo apenas os dados de m6, assim como antes, são de 1.095x200 (1095 observações em 200 componentes principais). Ainda, 750 observações foram separadas para treino e 345, para teste.

A representações em 3 dimensões dos dados de treino do microfone 6 e dos neurônios após treinamento estão na Figura 4.17. A representação dos neurônios em matriz U está em seguida, na Figura 4.18.

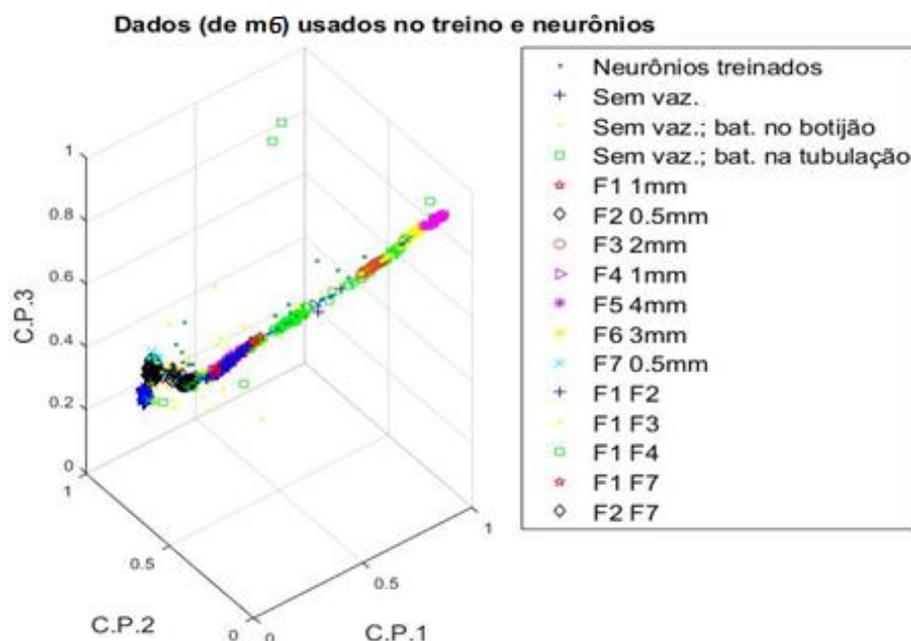


Figura 4.17 Neurônios treinados e os dados de treinamento padronizados de m6, em termos de suas componentes principais (C. P.)

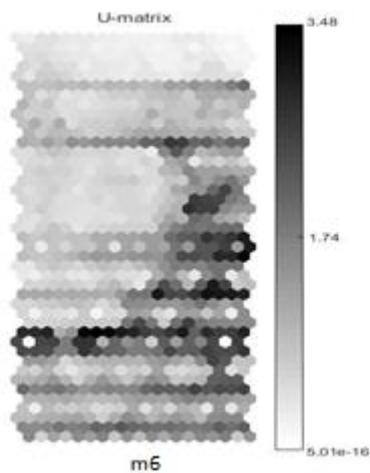


Figura 4.18 Representação dos neurônios em matriz U com dados apenas de m6

Os neurônios ativados por dados de treino e de teste estão evidenciados na Figura 4.19. Como anteriormente, nas figuras de (a) a (d) são calculados os neurônios ativos em cada configuração pelos dados de treino, enquanto em (e) e (f) são os neurônios ativados em cada configuração pelos dados de teste.

Ao se analisar as figuras (b) e (d) da Figura 4.19, verifica-se que o treinamento com dados do microfone 6 também possui distinções de área ativadas por dados de mesma configuração, ou configurações semelhantes.

As respostas de treino, (b) e (d), e as de teste (f) possuem muita similaridade na distribuição das cores e nos agrupamentos formados, o que demonstra confiabilidade na rede treinada.

No entanto, percebe-se que neste treinamento, já não há qualquer distinção entre os dados de vazamento em F1, em F4, em F1 e F2 simultâneos e em F1 e F7 simultâneos, grupo com o retângulo verde em seu interior na figura (b). Ou seja, todas as configurações de vazamento com 1 m de diâmetro e de vazamentos simultâneos de 0,5 mm e 1 mm não são distinguíveis por esta rede baseada nos dados de m6.

De modo semelhante, vazamentos em F3 (2 mm de diâmetro) e em F3 e F1 (2 e 1 mm de diâmetro) simultaneamente também são atribuídos ao mesmo grupo, evidenciado com círculo verde em seu interior na figura (b). A hipótese de que as amplitudes sonoras de um vazamento menor mais distante podem ser mascaradas por um vazamento maior mais próximo do microfone se encaixa também neste caso, visto que o microfone 6 se encontra mais próximo de F3 (2 mm) a 17,6 m do que de F1, a 19,1 m de distância.

Algo que se pode perceber também é que os grupos em amarelo (vazamento em F5) e em ciano (em F6), apesar de serem separados dois grupos distintos, possuem padrões próximos. Fato este inferido por haver hexágonos de coloração amarela no interior de hexágonos maiores de coloração ciano (mais vezes ativados por dados de vazamento em F6) na figura (b). Este resultado pode ser complementado com a análise feita da Figura 4.4 e com a da Figura 4.5 em que o padrão da resposta de vazamento em F5 era distinto dos demais padrões, porém, os padrões distintos eram em sua maioria respostas do microfone 1 e do microfone 2, enquanto que as resposta dos microfones mais distantes de F5 (m6, m11 e 13) não eram tão distintos dos padrões de outros vazamentos.

As configurações de vazamentos separadas por grupos distintos podem ser associadas a áreas de neurônios. Assim, de acordo com a área ativada na matriz U deste treinamento de dados de m6 identificam-se as seguintes configurações: área de neurônios ativados por condições sem vazamentos (preto, rosa, escuro e azul escuro), área por vazamentos com 0,5 mm de diâmetro (verde e cinza), dois vazamentos de 0,5 mm (azul celeste), vazamento com 1 mm ou- uma limitação da rede- com dois

vazamentos de 1 mm e de 0,5 mm (área com o retângulo verde em seu interior, na figura (b)), área de dois vazamentos de 1mm (em roxo), área de vazamento com 2 mm ou- outra limitação- dois vazamentos simultâneos de 2 mm e de 1 mm (área com o círculo verde em seu interior, na figura (b)), área de vazamento com vazamento de 3 mm (área em ciano) e área de vazamentos de 4 mm (área amarela).

Assim, de acordo com a área ativada, também, se identifica o sistema, porém com limitações: a rede por vezes confunde repostas de dois vazamentos simultâneos com a resposta apenas do maior vazamento. Desta forma, a rede identifica primeiramente o maior vazamento e, apenas depois deste consertado, passa a acusar o menor vazamento- não acusado anteriormente- e ainda em ocorrência.

5) Construção da Rede 5: Dados do Microfone 11

Apenas os dados do microfone 11, m11, para o treino da Rede 5 foram considerados. Este microfone se encontra fisicamente após os furos de vazamentos simulados, a uma distância de 33 m de F7, o vazamento mais próximo e a de 38,4 m de F1, o vazamento mais distante. A Rede 5 tem o objetivo de distinguir as perturbações ocorridas ao sistema.

O tamanho da matriz contendo apenas dados de m11 teve dimensões de 1.095x200 (1095 observações em 200 componentes principais). Destas observações, 750 observações foram separadas para treino e 345, para teste.

A representações em 3 dimensões dos dados de treino do microfone 11, evidenciados em configurações experimentais simuladas, e dos neurônios após treinamento estão na Figura 4.20.

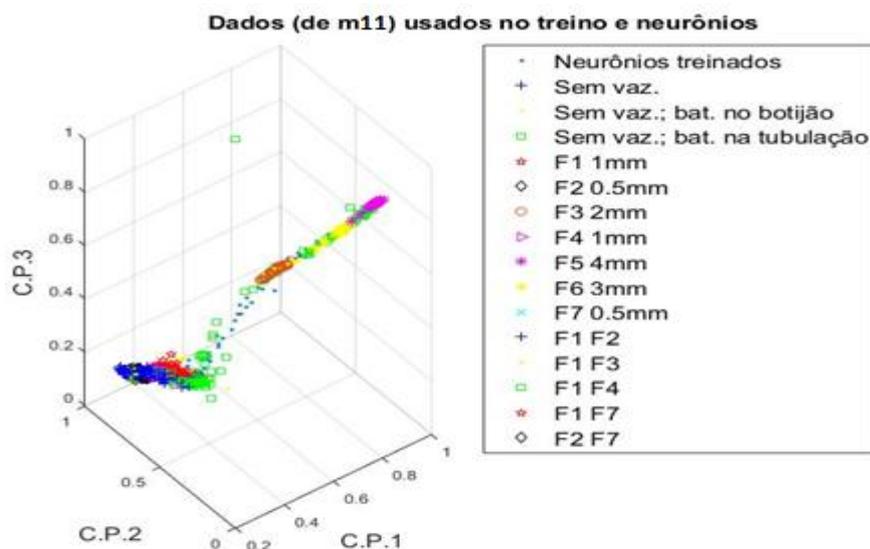


Figura 4.20 Neurônios treinados e os dados de treinamento padronizados de m11, em termos de suas componentes principais (C. P.)

A representação dos neurônios em matriz U, calculada levando em consideração todas as dimensões dos neurônios, está na Figura 4.21.

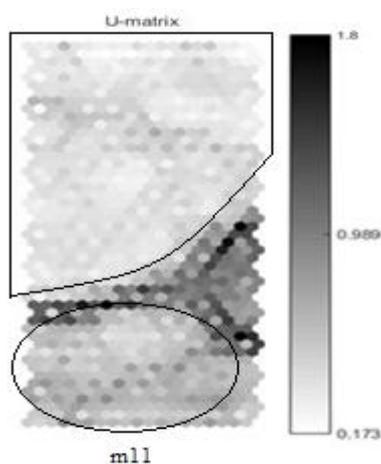


Figura 4.21 Representação dos neurônios em matriz U com dados apenas de m11

Na matriz U acima, há duas regiões mais claras evidenciadas. Isto significa que os neurônios representados nestas duas regiões se encontram próximos de seus neurônios adjacentes. Assim, estão associados a padrões com semelhanças.

Evidenciando quais neurônios são ativados por quais configurações, tem-se a Figura 4.22. De (a) a (d), são investigadas as ativações de neurônios pelos dados de treinamento. Enquanto que (e) e (f), dados de teste.

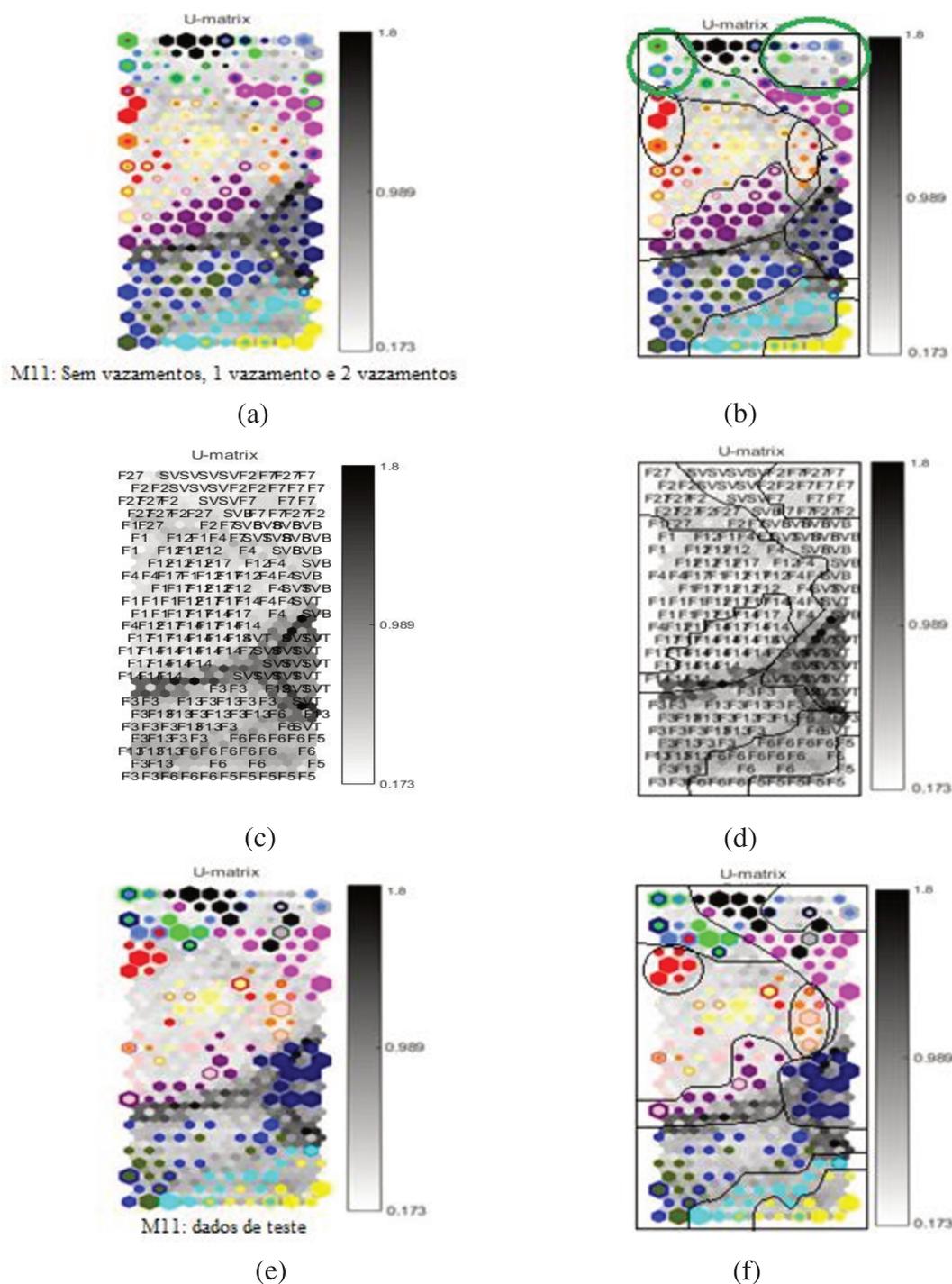


Figura 4.22 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27)

Pelos agrupamentos distinguidos nas figuras (b) e (d), o resultado da matriz U relativa aos dados de m11 também possui neurônios ativados por padrões específicos de configurações do sistema estudadas.

Percebe-se, na figura (b), nas duas áreas circuladas em verde, que os hexágonos em azul celeste (ativados por vazamentos simultâneos de F2 e F7, ambos de 0,5 mm de diâmetro) estão agrupados juntamente com hexágonos representando neurônios ativados majoritariamente por dados de vazamento de F2 e de F7 isolados. Assim, na rede treinada a partir dos dados apenas de m11 não se distingue dois vazamentos de 0,5 mm simultâneos de apenas um vazamento de 0,5 mm.

Há uma área que compreende os vazamentos de F1, de F4, de F1 e F2 simultâneos e de F1 e F7 simultâneos, ou seja, não foram distinguidos grupos para vazamentos de 1 mm e de dois vazamentos simultâneos de 1 e de 0,5 mm. Porém, percebe-se que há duas regiões (circuladas em preto dentro desta área, figura (b)) onde há uma tendência de apenas dados de 1 vazamento ativarem os neurônios.

Como na rede anterior, relativa a m6, não há diferenciação de uma área ativada por vazamentos de F3 (de 2 mm) e vazamentos simultâneos de F3 e F1 (2 e 1 mm), sendo duas situações que se confundem, representada pela área com hexágonos em azul e em verde escuro na parte de baixo da matriz U, figuras (b), (d) e a de teste (f).

De todo modo, as áreas diferenciadas representam neurônios que são ativados por uma ou mais configurações, que devem ser similares, de vazamentos. Assim, novamente, de acordo com a área ativada na matriz U deste treinamento de dados de m11, identificam-se estas configurações: área de neurônios ativados por em condição de sem vazamentos (hexágonos em preto, rosa, escuro e azul escuro), por vazamentos com 0,5 mm de diâmetro ou- primeira limitação da rede- de dois vazamentos de 0,5 mm (verde, cinza e azul celeste), por vazamentos com 1 mm ou- segunda limitação- com dois vazamentos de 1 mm e de 0,5 mm (áreas circuladas em verde na figura (b)), por dois vazamentos de 1mm (em roxo), por vazamento com 2 mm ou- terceira limitação- dois vazamentos simultâneos de 2 mm e de 1 mm (hexágonos em verde escuro e azuis), por vazamento com de 3 mm (hexágonos em ciano) e, finalmente, por vazamentos de 4 mm (hexágonos em amarela).

De acordo com a área ativada, com as limitações mencionadas acima, se identifica o sistema. Novamente, como na Rede 4, observou-se que, em casos que ocorram mais de um vazamento, os neurônios ativados são os mesmo que ativam na ocorrência apenas do maior vazamento. Assim, apenas corrigindo o maior, o menor

começa a ser identificado pela rede. Na Rede 5, acrescenta-se o caso de que quando dois vazamentos de 0,5 mm de diâmetro ocorrem, os neurônios ativados são os mesmos neurônios ativados quando há apenas um vazamento de 0,5 mm. Assim, nesta situação, caso se corrija um vazamento, a rede continuaria acusando vazamento de 0,5 mm pela existência de um segundo.

6) Construção da Rede 6: Dados do Microfone 13

Apenas os dados do microfone 13 foram considerados, m13, para o treinamento da Rede 6. Fisicamente, no sentido do fluxo do ar comprimido, este microfone se encontra após os furos de vazamentos simulados, a uma distância de 40,1 m de F7, o vazamento mais próximo e a de 45,5 m de F1, o vazamento mais distante. A Rede 6 tem o objetivo de distinguir as perturbações ocorridas ao sistema.

Os dados de m13, para o treinamento da rede, possuía 750 observações de amplitude em termos de componentes principais (200 componentes) e, para teste 345. Todas as matrizes contendo apenas um microfone possuem estas configurações.

A representações em 3 dimensões dos dados de treino do microfone 13, evidenciando as diferentes configurações experimentais, e dos neurônios após treinamento estão na Figura 4.23.

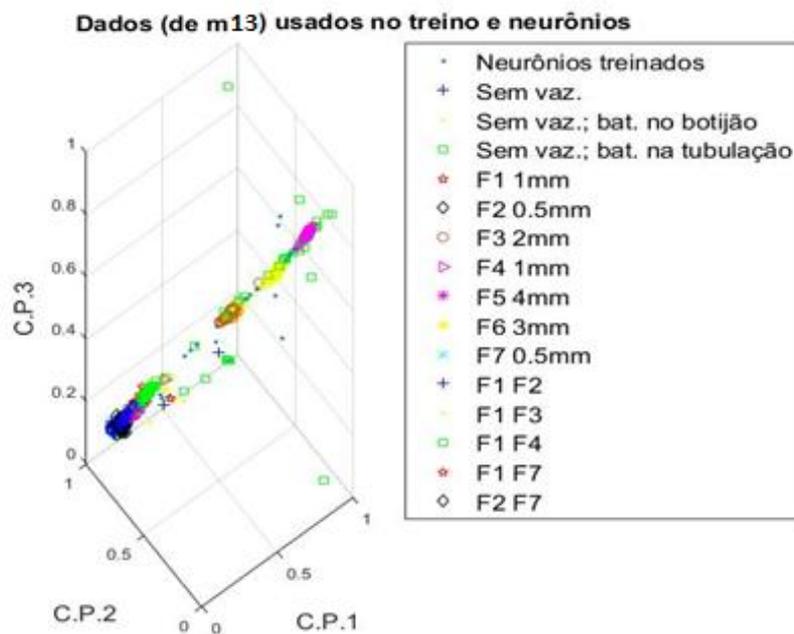


Figura 4.23 Neurônios treinados e os dados de treinamento padronizados de m13, em termos de suas componentes principais (C. P.)

A representação dos neurônios em matriz U é mostrada na Figura 4.24.

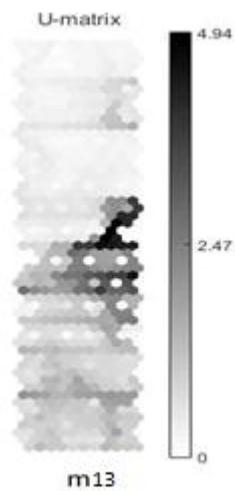


Figura 4.24 Representação dos neurônios em matriz U com dados apenas de m13

Os hexágonos de coloração clara, de acordo com a escala ao lado da matriz U, significa que os neurônios representados pelo respectivo hexágono se encontram perto dos vizinhos adjacentes.

A Figura 4.25 evidencia os neurônios ativados por dados de treino e teste. De (a) a (d) são os calculados os neurônios ativos em cada experimento pelos dados de treino enquanto em (e) e (f) são os neurônios ativados em cada experimento pelos dados de teste.

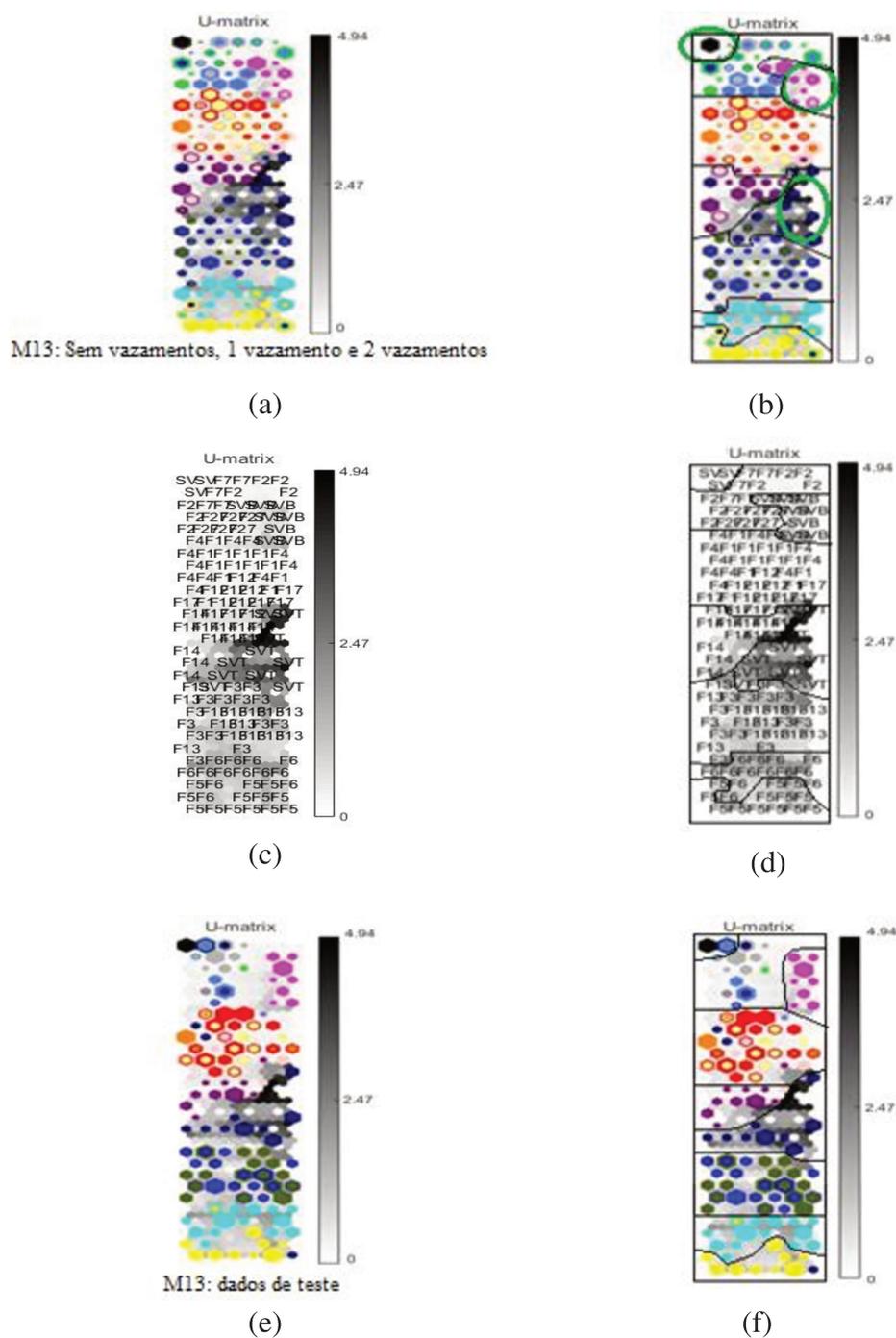


Figura 4.25 Representação na matriz U dos neurônios ativados por dados de treino ((a)-(d)) e dados de teste ((e) e (f)). À direita, agrupamentos visuais de configurações experimentais similares. Situações sem vazamento- preto: sem vazamento (SV); rosa escuro: batidas no botijão (SVB); Azul escuro: batidas na tubulação (SVT); Com vazamentos- vermelho: em F1 (F1); verde: em F2 (F2); azul: em F3 (F3); laranja: em F4 (F4); amarelo: em F5 (F5); ciano: em F6 (F6); cinza: em F7 (F7); amarelo claro: em F1 e F2 (F12); verde escuro: em F1 e F3 (F13); roxo: em F1 e F4 (F14); rosa claro: em F1 e F7 (F17); azul celeste: em F2 e F7 (F27)

Comparando-se as distribuições de cores entre as figuras (b) e (f) acima, verifica-se que são semelhantes. Assim, qualitativamente, o teste está de acordo com o treino. Ou seja, considera-se que foi realizado um bom treinamento da rede neural.

Quanto as áreas distinguidas para mesmos experimentos, ou experimentos similares, os resultados de m13 não diferem muito daquele de m11. Uma diferença, porém não relevante para caracterizar vazamentos, foi a atribuição de áreas diferentes às configurações com ausência de vazamentos, são as áreas com círculo verde em seu interior.

7) Conclusão sobre as Redes 2 a 6

Em resumo, a Rede 2 (com os dados apenas de m1) e a Rede 3 (com dados apenas de m2) conseguiram distinguir áreas para todas as configurações diferentes estudadas. Porém, no teste, a Rede 3 já apresentou dificuldades de dissociar vazamentos em F3 e em F1 e F3 simultaneamente.

Na Rede 4, com dados de m6, houve uma área de neurônios ativados por configurações de vazamentos em F1, F4, F1 e F2 e F1 e F7, o que implica que a rede já não diferencia bem vazamentos de 1 mm de dois vazamentos simultâneos de 1 mm e de 0,5 mm. Também não houve a diferenciação entre vazamento em F3 e em F1 e F3 simultaneamente, posto que dados das duas configurações ativam neurônios de uma mesma área. Esta área, assim como todas em que ocorrem estas limitações, não pode ser claramente distinta em duas, uma para cada configuração.

Enquanto que em Rede 5 e em Rede 6 mais uma configuração passou a não se distinta: vazamento em dois vazamentos simultâneos de 0,5 mm, passando a ativar os mesmos neurônios que os dados de apenas um furo de 0,5 mm.

Estas limitações não se mostram, em primeira vista, graves, posto que elas apenas não identificam que há dois vazamentos, mas identificam um de cada vez, o maior vazamento primeiro. A identificação ainda ocorre. Uma vez sanado o maior vazamento, o segundo seria imediatamente identificado.

8) Respostas de Interpolação e de Extrapolação na Rede de Kohonen

Para se testar as respostas da rede não supervisionada em situações não treinadas pela rede foram testados em dados apenas de m1.

A situação de interpolação foi testada com a retirada de vazamentos em F3 do treinamento da rede. Visto que o diâmetro de F3 é de 2 mm, enquanto que dados de vazamentos com 1 mm de diâmetro e com 3 mm de diâmetros estão presentes no treino, a situação passa a ser de interpolação.

Para retirar os vazamentos em F3, a situação de vazamentos simultâneos de F1 e F3 também foi retirada do treinamento.

Os dados foram padronizados por equalização de histograma e, em seguida, os dados de m1 foram separados para treinamento e teste da rede. Os dados de vazamentos de F3 isolado e de F1 e F3 simultâneos foram usados apenas para teste.

O resultado da matriz U está na Figura 4.26 e está evidenciando neurônios ativados pelos dados de treino (a) e de teste (b).

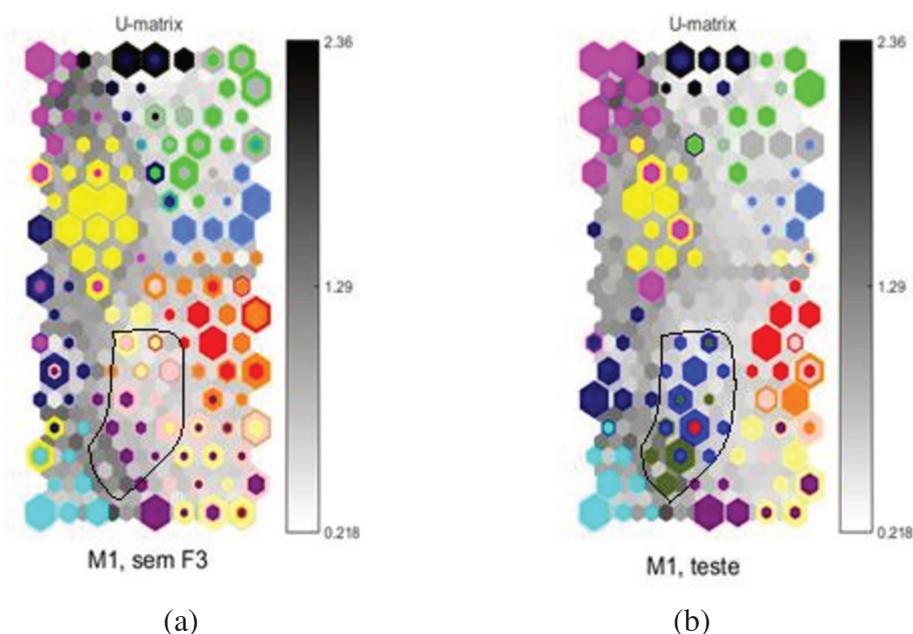


Figura 4.26 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com interpolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7

A área ativada pelos dados de vazamentos em F3 testados na rede (em azul, figura (b)) e pelos de vazamentos simultâneos em F1 e F3 (em verde escuro, figura (b))

está em destaque na figura (b) e, para fins de comparação, na figura (a): percebe-se que os dados de vazamento de F3 e em F1 e F3 simultâneos, ativaram neurônios que foram treinados para serem ativados por dados de situações de dois vazamentos de 1 mm (F1 e F4, em roxo) e de vazamentos simultâneos de 1 mm e 0,5 mm (F1 e F7, em rosa claro e F1 e F2, amarelo claro).

Com a ausência de dados de vazamento com 2 mm de diâmetro no treinamento da rede, não havia uma área própria para a identificação de suas características. Assim, os dados de teste correspondentes a este vazamento ativaram neurônios correspondentes a dados contidos no treinamento, no caso, de duas situações experimentais, que são as com maiores similaridades, as com padrões mais próximos aos dados de teste.

Em uma situação prática, pode-se perceber que a resposta acusada pela matriz U se trata de uma interpolação levando-se em conta também a resposta dos microfones ao longo do tempo, por exemplo. Pois, percebe-se que as respostas de F3 e de F1 e F3 simultâneos possuem respostas no tempo diferentes das configurações acusadas pela rede na interpolação, como visto no Apêndice C.

Outro teste investigado foi a situação de extrapolação da rede. A extrapolação foi testada com a retirada de vazamentos em F5 do treinamento da rede. Este vazamento possui o maior diâmetro em estudo, de 4 mm. Com a ausência deste vazamento no treinamento da rede, o maior diâmetro de treino é de 3 mm, correspondente ao vazamento em F6. Os dados de vazamentos de F5 foram usados apenas para teste.

O resultado da deste treinamento matriz U está na Figura 4.27 e, novamente, está evidenciando neurônios ativados pelos dados de treino (a) e de teste (b).

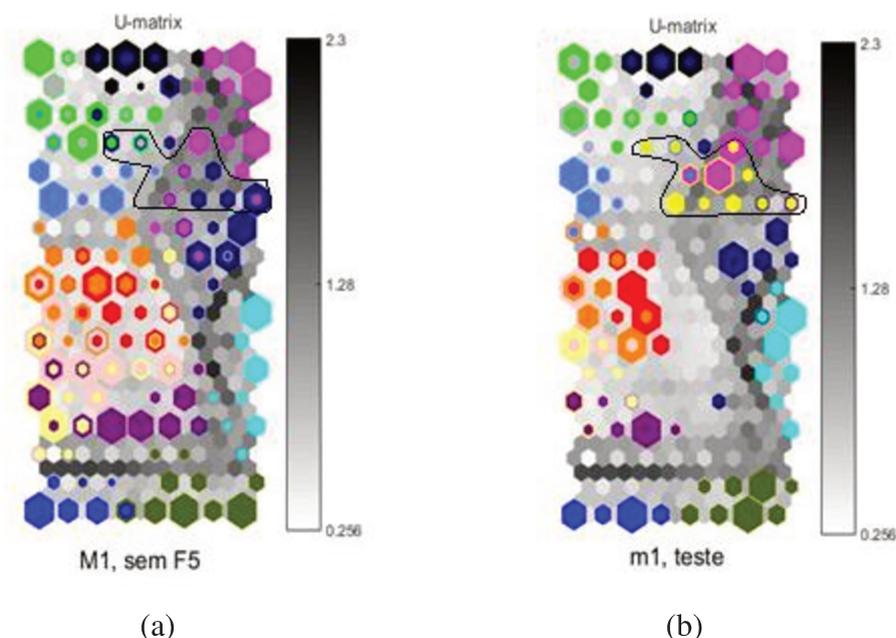


Figura 4.27 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com extrapolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7

A área ativada pelos dados de vazamentos em F5 testados na rede (em amarelo, figura (b)) está em destaque na figura (b) e na figura (a) para fins de comparação. Ao se analisar as respostas de treino e de teste compreendidas nas áreas em destaque, percebe-se que, os dados de vazamento em F5, de diâmetro de 4 mm, ativam neurônios associados principalmente a situações de sem vazamento com batidas no botijão (em rosa escuro, figura (a)) e na tubulação (em azul escuro, figura (a)).

Assim, nesta extrapolação, mostra-se que os dados de F5 possuem maiores similaridades com os dados de sem vazamento com batidas no botijão. A resposta da rede será sempre a configuração presente no treinamento com maior afinidade com o dado de teste. O erro na identificação do sistema seria facilmente perceptível neste caso, pois acusa ausência de vazamento quando há vazamento, qualquer outro indício de vazamento revelaria se tratar de um padrão falsamente reconhecido pela rede. O próprio comportamento da amplitude sonora no tempo poderia ser um indício, conforme visto no Apêndice C, pois as situações sem vazamento com batidas possuem comportamentos distintos daquela com vazamento em F5.

A Tabela 4.3 possui algumas características das redes para teste de interpolação e de extrapolação feitas. Todas foram feitas apenas com dados de m1, com a organização hexagonal em topologia plana. O algoritmo de treinamento de ambas foi o sequencial com a função “power” de taxa de aprendizado e inicialização linear dos pesos.

Tabela 4.3 Parâmetros de treinamento das redes de interpolação e de extrapolação

| Rede | Tamanho da Rede | E. Q. | E. T. | Função de Vizinhança | Objetivo da Rede |
|-----------------|-----------------|-------|-------|----------------------|---------------------|
| De interpolação | 17x8 | 3,198 | 0,003 | <i>Gaussian</i> | Testar interpolação |
| De extrapolação | 17x8 | 3,131 | 0 | <i>Gaussian</i> | Testar extrapolação |

Um outro tipo de extrapolação testada foi em relação à distância do vazamento, em vez de diâmetro de vazamento.

Assim, pegou-se a Rede 4, treinada e testada anteriormente com dados de m6, e se testou dados de m2. As mesmas configurações experimentais ocorrem, porém a distâncias diferentes.

A matriz U da Rede 4 está na Figura 4.28. Nela, estão evidenciados os neurônios ativados por dados capturados por outro microfone, o m2.

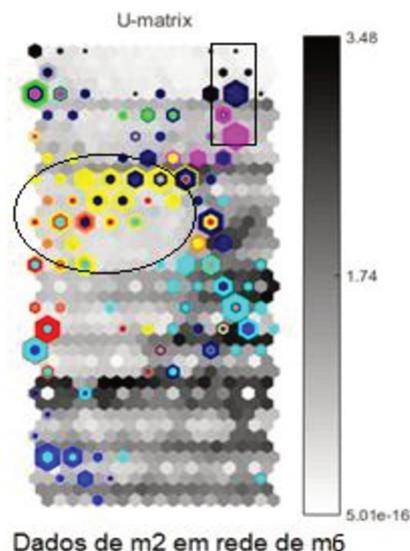


Figura 4.28 Representação na matriz U dos neurônios ativados por dados de treino (a) e dados de teste com extrapolação (b). Situações sem vazamento- preto: sem vazamento; rosa escuro: batidas no botijão; Azul escuro: batidas na tubulação; Com vazamentos- vermelho: em F1; verde: em F2; azul: em F3; laranja: em F4; amarelo: em F5; ciano: em F6; cinza: em F7; amarelo claro: em F1 e F2; verde escuro: em F1 e F3; roxo: em F1 e F4; rosa claro: em F1 e F7; azul celeste: em F2 e F7

O resultado do teste da figura acima se mostra com algumas concordâncias em comparação ao resultado das ativações de dados do treinamento da rede, Figura 4.19 (b), por exemplo, a região no retângulo acusando ausência de vazamento.

No entanto, em relação à dimensão do vazamento, em sua maioria, não há muitas concordâncias. A exemplo, a região circulada de hexágonos em amarelo é ativada por dados de teste de m2 de vazamento em F5, de 4 mm de diâmetro. Nos dados de treino desta rede, essa região foi associada a vazamentos de 1 mm, ou de 1 mm juntamente com 0,5 mm (F1, F4, F1 e F2 e F1 e F7).

Dessa forma, mostra-se que as redes treinadas a partir de dados de um microfone são capazes de identificar o sistema, algumas com limitações, anteriormente mencionadas. Porém, para se testar a rede, é importante que os dados de teste, para uma resposta mais assertiva, tenham uma distância em relação ao microfone próxima daquela de dados de treino.

A principal conclusão deste último teste de extrapolação é a de mostrar ser possível, mesmo a uma distância entre o vazamento e o microfone diferente daquela de treino, dentro da área de alcance do sinal sonoro, a técnica proposta ainda pode

identificar a região de padrão “sem vazamento” para “evento anormal”, o que poderia soar um alarme indicando a anormalidade. O Apêndice E mostra outras extrapolações em relação à distância.

Na seção a seguir, foram analisadas redes treinadas com a ausência de compressão dos dados, sem PCA.

4.5.2 Testes de Redes Não Supervisionadas Sem PCA

Em testes de alternativas para a compressão de dados sem a utilização da técnica de análise de componentes principais (PCA) e para sua possível avaliação, foram realizados os testes com dados diretamente no domínio da frequência para o treinamento das redes com dois esquemas para redução da matriz de dados: escolhendo aleatoriamente amplitudes de respectivas frequências, e usando amplitudes de todas as frequências, porém apenas em regime permanente.

1) Com Amplitudes em Frequências Aleatórias

Se a matriz de dimensões 5.475x6.599, contendo todos os dados de todos os microfones, fosse diretamente usada como dados de treino para a rede não supervisionada de Kohonen, um grande esforço computacional seria exigido.

Para diminuição deste esforço, sem qualquer tratamento para diminuição da dimensionalidade, foi aleatoriamente escolhido apenas os 1.000 primeiros pontos de frequência para o treino. Além disso, apenas dados do microfone 1, m1, foram usados neste teste. Ou seja, uma amostra da matriz foi utilizada, cuja dimensão foi de 1.095x1.000 (1.095 observações do microfone 1 de amplitudes em cada uma das 1.000 primeiras frequências em que amplitudes foram calculadas por FFT, que eram dentro do intervalo de 100 a 1.600 Hz). Na Figura 4.29, na Figura 4.30 e na Figura 4.31, a representação das cores é a mesma dos gráficos anteriores das matrizes U.

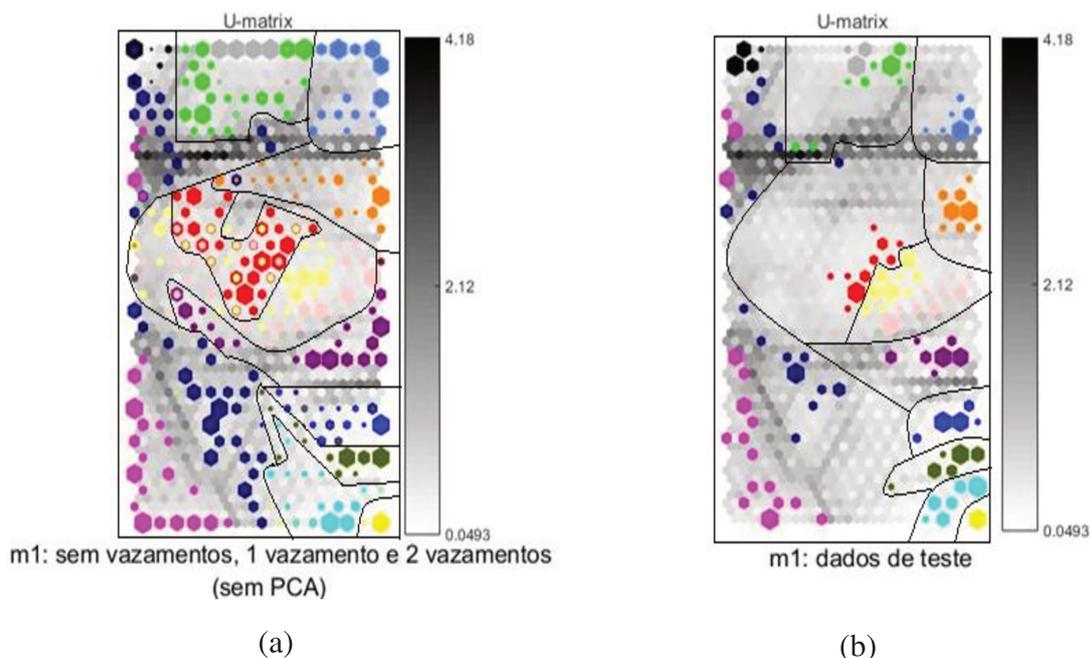


Figura 4.29 Matriz U com dados de m1 com amplitudes em frequências de 100 a 1.600 Hz

Em comparação com a matriz U exposta na Figura 4.13 treinada com dados após PCA, também com dados de m1, a separação de áreas de neurônios responsáveis por configurações diferentes de experimento foi facilitada com a ausência da compressão dos dados, percebe-se que as colorações diferentes possuem áreas melhores definidas. No entanto, a escolha das frequências a serem utilizadas foi completamente arbitrária.

Na Figura 4.29, observa-se ser possível utilizar a rede de Kohonen sem a compressão dos dados (por PCA). Porém, sem a compressão, seria necessária uma escolha mais precisa das frequências dominantes para se evitar escolher dados para o treinamento da rede que não sejam representativos do sistema.

Os dados para treino da rede na Figura 4.29 foram primeiramente separados da matriz com todos os dados (observações em frequência). Em seguida, normalizados por equalização de histograma e, após, apenas os dados de m1 foram considerados. A rede treinada possui 450 neurônios (30x15), vizinhança hexagonal, topologia plana, com inicialização linear, algoritmo de treinamento em batelada, função de vizinhança gaussiana, erro de quantização de 2,426 e erro topográfico de 0.0013.

O mesmo teste foi conduzido com o microfone mais distante, m13. Novamente, a não compressão dos dados facilita a associação de agrupamentos de neurônios ativados por uma mesma configuração. Os agrupamentos de vazamentos que envolvem

F1 (em vermelho), F4 (em laranja), F1 e F2 (em amarelo claro), F1 e F7 (em rosa claro) são distintos (regiões evidenciadas entre as duas retas), o que não ocorre com a mesma análise com os dados em componentes principais na Figura 4.25.

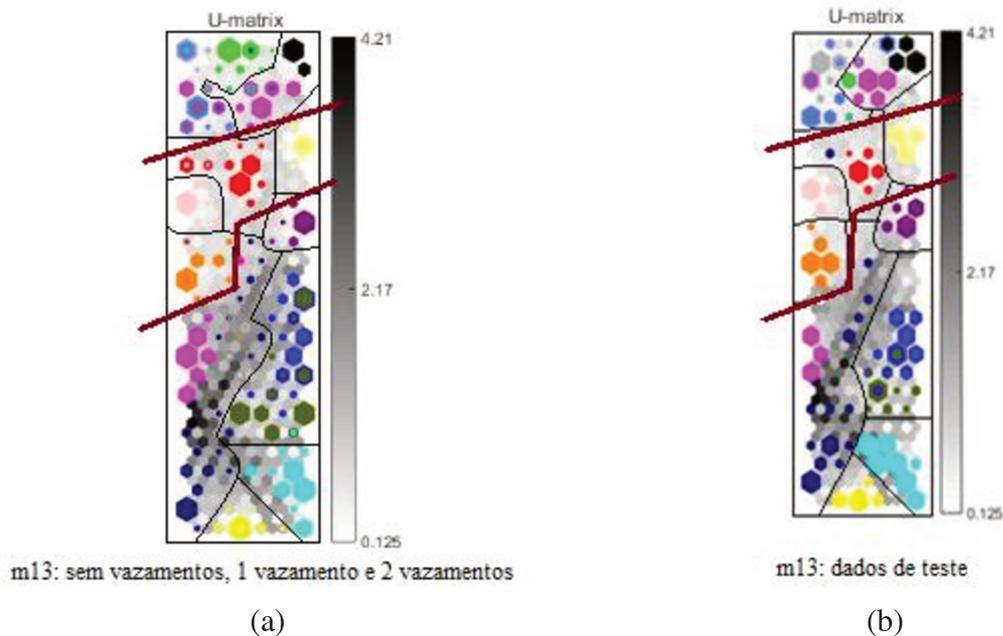


Figura 4.30 Matriz U com dados de m13 com amplitudes em frequências de 100 a 1.600 Hz (a) Evidenciando neurônios ativados pelos dados de treino; (b) Evidenciando neurônios ativados pelos dados de teste

Novamente, na Figura 4.30, as cores possuem o mesmo sentido que nas matrizes U anteriores. Os dados de treinamento da rede na Figura 4.30 foram primeiramente separados da matriz com todas as demais frequências em estudo. Em seguida, normalizados por equalização de histograma e, então, apenas dados de m13 foram considerados. A rede treinada possui 156 neurônios (26x6) com organização hexagonal da vizinhança e topologia plana. A inicialização dos pesos foi a linear, o algoritmo de treinamento utilizado foi o em batelada e a função de vizinhança foi a gaussiana. Erro de quantização de 3,4332 e erro topográfico de 0.0053.

Assim, o uso da rede pode ser realizado com os dados em termos de frequência, mas o método seria inviável com o aumento de configurações ou de tempo de captação de dados. Além, a falta de um método para a escolha das frequências faz com que a amostragem seja potencialmente não representativa do sistema e excessiva.

Faz-se necessário um estudo para a escolha das frequências principais para esta análise.

2) Com Amplitudes em Regime Permanente

O objetivo deste teste foi de investigar uma possível diminuição da quantidade de dados de treinamento sem excluir frequências em análise, como feito no teste anterior. Assim, também não há a necessidade da compressão de dados realizada na análise de componentes principais (PCA).

Apenas dados com aproximadamente 60 s foram utilizados. Neste tempo, julga-se que as amplitudes estejam em regime permanente, ou próximo disso.

Assim, novamente uma amostra da matriz de dimensões 5.475x6.599 foi utilizada: separaram-se apenas os últimos dados gerados por m1, com aproximadamente 60 s. A matriz resultante para treino teve dimensão de 60x6.599 (60 observações de amplitudes em 6.599 frequências diferentes, no intervalo de 100 Hz a 10.000 Hz). Enquanto a matriz resultante de teste teve dimensões de 45x6.599.

Os dados usados nesta rede foram também normalizados por equalização de histograma. Desta vez, para a normalização, apenas os dados já separados para treino e para teste foram considerados. A rede treinada possui tamanho do mapa de neurônio de 20x9. A organização da vizinhança foi hexagonal e topologia plana. Foram empregados para treinamento da rede, inicialização linear dos pesos, algoritmo batelada de treinamento e função de vizinhança gaussiana. O erro de quantização resultante foi de 5,4584 e o erro topográfico, 0.

A matriz resultante do treino está representada juntamente com o teste na Figura 4.31.

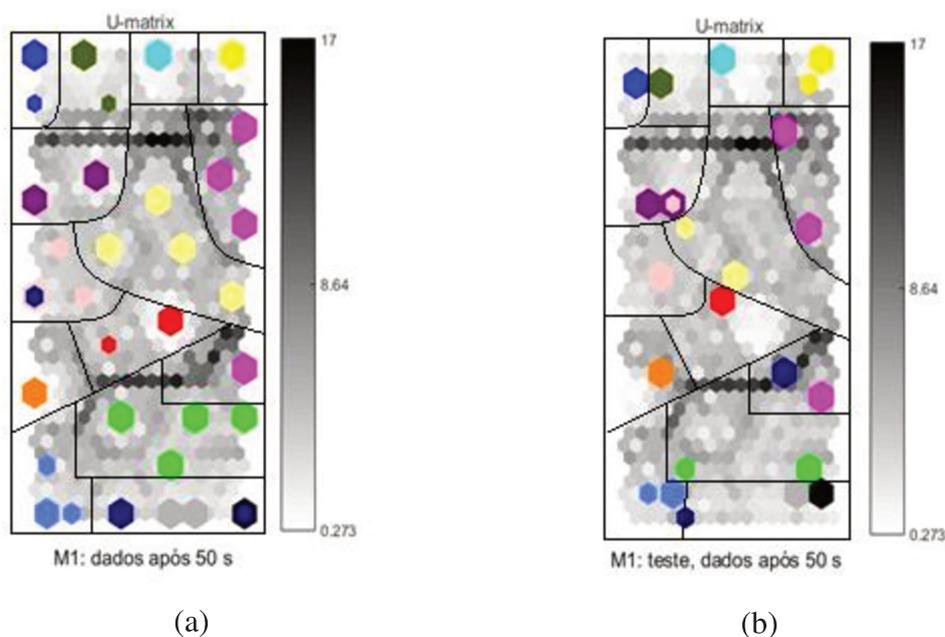


Figura 4.31 Matriz U com dados de m1 com amplitudes em tempo de simulação de aproximadamente 60 s. (a) Evidenciando neurônios ativados pelos dados de treino; (b) Evidenciando neurônios ativados pelos dados de teste

Na Figura 4.31, as cores possuem o mesmo sentido que nas matrizes U anteriores. Verifica-se uma distribuição de cores parecidas nos neurônios ativados por dados de treino e de teste, o que pode indicar que a metodologia também possa ser utilizada apenas para dados em regime permanente, o que diminuiria a dimensão de dados, caso se aceite que o vazamento não será caracterizado enquanto estiver em seu regime transiente.

Porém, não há muitas concordâncias exatas nos neurônios ativados. A variação de neurônios ativados, diferenças na resposta de um instante ao outro mesmo em regime permanente, pode ser causada por perturbações não controladas ao qual o sistema possa estar sujeito.

Assim, um trabalho futuro poderá investigar a criação e a efetividade de redes não supervisionadas com a obtenção apenas de dados em regime permanente.

De uma forma ou de outra, sugere-se que a rede treinada apenas com os dados de vazamentos em regime permanente para um futuro monitoramento *on-line*, na tentativa de se diminuir a base de dados de treino, ainda possua dados suficientes de forma que a rede gerada não seja suscetível demais a erros por perturbações não controladas mesmo com o mínimo possível de perturbações externas.

4.6 Propostas de Uso em Montagem Física em Sistema Equivalente

Pelos resultados qualitativos obtidos, aceitando que até a distância de 19 m, de 19 m ao vazamento mais distante, se tem a identificação do sistema com poucas limitações. Propõem-se, então, em tubulação equivalente à estudada neste trabalho, acoplamentos de microfones espaçados de 38 em 38 m, deixando cada microfone responsável pela identificação da área 19 m a frente e 19 m atrás de sua posição.

O microfone mais perto da perturbação seria identificado como aquele com a maior variação em sua amplitude no domínio do tempo.

Em sua área de ação, a rede deve conter dados de treino de configurações diferentes de vazamentos e de batidas espaçadas de -19 m a +19 m do local do microfone. Assim, primeiramente, para determinar a posição da perturbação, redes como a Rede 1 podem ser utilizadas; em seguida, para identificar a perturbação, redes treinadas com dados a uma distância específica do microfone podem ser usadas, como as redes 2 a 6 deste trabalho.

Para aumentar a área de ação, outros microfones podem ser testados, ou se explorar outros métodos de se diminuir o tamanho da matriz de dados.

Os testes sem PCA feitos sugerem também outras pesquisas. Por exemplo, o microfone m13 (com 45,5 m de distância do vazamento mais longe), com apenas as amplitudes em frequências de 100 Hz a 1.600 Hz, teve suas áreas na matriz U definidas com menos limitações do que quando o PCA foi utilizado. Assim, o alcance do microfone pode ser estendido a mais do que 19 m com outros tratamentos dos dados.

4.7 Sugestões para Trabalhos Futuros

Neste trabalho, apenas o sistema inicialmente em 1 bar foi investigado. Porém, é de se esperar que a pressão interna do sistema em vazamento influencie a resposta sonora. Como em situações práticas, a pressão de trabalho de uma tubulação pertence à uma faixa de valores, avaliar como o acréscimo de dados, observações em pressões próximas à pressão principal da investigação, afetaria na identificação do sistema é uma tarefa futura.

Para processamento do sinal, uma sugestão é o uso de Transformadas de *Wavelet*. Esta decompõe o sinal tanto no domínio da frequência quanto do tempo, podendo-se correlacionar respostas de frequência diretamente com o tempo.

Diferentemente do resultado da FFT que não gera nenhum resultado sobre o tempo em que a resposta do sistema teve as amplitudes nas frequências calculadas (assim, neste trabalho, esta correlação das amplitudes em frequência com o tempo teve que ser realizada de acordo com a aquisição de dados conforme exposto na seção 4.1 deste trabalho).

Como já sugerido, outros métodos de compressão dos dados, ou de métodos para realizar escolhas significativas amplitude em termos de frequência, podem ser investigados no intuito de facilitar e deixar mais eficiente a associação de neurônios aos padrões correspondentes, como visto nos testes sem PCA.

Outra sugestão é a realização da investigação *on-line* do vazamento, posto que o presente trabalho foi *on-line* apenas enquanto uso do software *Labview*.

As análises das redes se deram de forma qualitativa e de modo visual atreladas à matriz U. Porém, uma análise quantitativa pode ser muito bem realizada, verificando-se numericamente quais neurônios são ativados e quantas vezes cada um é ativado. Assim, uma análise numérica, envolvendo análises estatísticas, para se associar os neurônios aos padrões que mapeiam e correlações para problemas de interpolações podem ser investigadas. A matriz U pode até mesmo não ser utilizada em análises quantitativas, posto que é uma ferramenta muito direcionada à visualização.

5 CONCLUSÕES

A partir dos dados obtidos no domínio do tempo, a estimativa direta da localização de um vazamento a partir dos dados de variação de amplitude sonora de apenas um microfone não se mostrou eficiente, pois, como visto, a resposta de variação de um vazamento de 0,5 mm (F2) de diâmetro captada por um microfone próximo (m1) foi similar àquela de um vazamento de 1 mm captada por um microfone mais distante (m13). Assim, a amplitude da onda sonora se mostrou dependente não só da localização, mas da dimensão do vazamento.

No entanto, ao se comparar respostas de variação de mais de um microfone, foi verificado que independentemente da dimensão do vazamento, a maior variação de amplitude da onda sonora será acusada pelo microfone mais próximo ao vazamento. Assim, pode-se determinar qual o microfone mais próximo do vazamento de modo rápido.

Como os dados adquiridos incluem o comportamento acústico desde o início do vazamento e ao longo de 50 s após início da perturbação (tempo considerado suficiente para alcançar regime permanente nas simulações estudadas), estes, portanto, expressam o comportamento acústico do sistema no regime transiente e no regime permanente. Sendo assim, a rede é treinada de modo a identificar a ocorrência de seu vazamento desde o momento que começa a ocorrer.

A análise de FFT gera dados no domínio da frequência, estes dados mostraram-se representativos do sistema para a identificação e para a caracterização de vazamentos. Esta conclusão é feita a partir tanto dos resultados dos testes de redes de Kohonen diretamente com dados resultados da FFT quanto das redes com a compressão de dados por PCA, visto que o resultado do PCA era originariamente em termos de frequência.

A realização do método de Análise de Componentes Principais (PCA) permitiu a compressão dos dados com observações em 6.599 dimensões para 200 componentes principais com 98,19% de representatividade, ou de variância original.

Porém, primeiramente, pôde-se fazer uma análise preliminar dos padrões a partir dos dados comprimidos a 3 dimensões, representando apenas 47,26% da variância original. A partir desta análise visual preliminar, percebeu-se que dados de uma mesma situação experimental de perturbação ao sistema tendiam a se agrupar, porém, que um

mesmo experimento podia possuir mais de um agrupamento. Estes agrupamentos diferentes foram atribuídos a respostas de microfones diferentes, assim, não só a configuração experimental possuía um padrão próprio, como cada microfone (posições diferentes na tubulação) também.

A partir destes agrupamentos distintos de padrões, mostrou-se ser possível distinguir não só o tipo de vazamento (ou de situações sem vazamento) como também determinar aproximadamente a distância do vazamento ao microfone. A partir da análise preliminar, percebeu-se que o mapeamento de neurônios poderia identificar a posição e a dimensão do vazamento, posto que cada situação aparentava ter um padrão próprio.

Em seguida, após a análise preliminar, as redes foram treinadas considerando a compressão dos dados do domínio da frequência a 200 componentes principais. O aumento do número de componentes principais utilizadas na compressão, acima de 200, não representa um aumento significativo na representatividade; por exemplo, o aumento de 50 dimensões, usando 250 componentes principais, há uma representatividade de 98,89%: um aumento de 0,7% em 50 dimensões (uma taxa de 0,014%/dimensão), não justificando o esforço computacional a mais que se teria nesse aumento de dimensões.

As análises das redes criadas foram qualitativas, os neurônios foram visualmente associados na matriz U a configurações experimentais. Em todas as redes, foi possível o agrupamento de neurônios respectivos a um mesmo grupo.

Na Rede 1, dados de todas as perturbações experimentais de todos os 5 microfones foram utilizados no treinamento. Como resultado, a rede foi capaz de associar neurônios a distâncias do microfone ao vazamento. Confirmando a análise preliminar de que dados a distâncias diferentes possuíam padrões diferentes. Nesta rede, também existe a possibilidade de se analisar a dimensão do vazamento em questão, ou a ausência de vazamentos, mas por se tratar de uma rede muito grande (compreende todos os dados em análises), para evitar perdas de informações mais detalhadas, optou-se por utilizar esta rede apenas para determinar a distância aproximada do vazamento em relação ao microfone.

Assim, as Redes 2 a 6 foram treinadas com dados apenas de um microfone: o objetivo de cada uma era de caracterizar o tipo de perturbação no sistema uma vez que a distância já tenha sido estabelecida. Cada uma destas redes mostrou a capacidade de diferenciar dimensões de vazamentos de outras e quanto mais próximo o microfone dos vazamentos, mais eficiente esta diferenciação.

A Rede 2, com dados de m1, microfone próximo aos vazamentos, diferenciou todas as configurações de dimensões de vazamento: áreas de neurônios para o reconhecimento de vazamento de 0,5 mm, de 1 mm, de 2 mm, de 3 mm, de 4 mm e áreas respectivas para dois vazamentos simultâneos como dois vazamentos de 0,5, de 1mm e 0,5mm, de ambos com 1 mm e de 1 mm e 2 mm (todas as configurações de diâmetros de vazamentos realizadas). Além, também houve áreas respectivas a situação de ausência vazamentos (com batidas ou não).

A Rede 3, com dados de m2, também próximo aos vazamentos, teve resultado similar ao da Rede 2: áreas de neurônios distinguidas com mesmos significados. Porém, nesta rede já houve dificuldade em distinguir dados de vazamento em F3 e de vazamentos simultâneos de F1 e F3. A explicação para este resultado é de que o maior vazamento (F3, 2 mm) está mais próximo ao microfone 2 do que o menor vazamento (F1, 1 mm), fazendo com que, em vazamentos simultâneos, o maior vazamento prevaleça sobre o menor de forma que as respostas acústicas sejam similares àquelas de apenas vazamento no maior furo (F3).

A Rede 4, com dados de m6, microfone a quase 20 m do vazamento mais distante, teve os resultados de vazamentos em F3 e em F1 e F3 simultâneos completamente indistinguíveis. Além disso, vazamentos de 1 mm e de 1 mm e 0,5 mm simultâneos também não foram separáveis, ou seja, vazamento em F1, em F4, em F1 e F2 simultâneos e em F1 e F7 simultâneos não são diferenciados por esta rede. Deste modo, apenas ao estancar o maior vazamento, o menor passaria a ser identificado pela rede.

Na Rede 5, com dados de m11, já a quase 40 m de distância do vazamento mais distante, mais uma região passa a não possuir área própria de neurônios: vazamentos simultâneos de 0,5 mm (F2 e F7). Estes passam a ser identificados pelos mesmos neurônios que identificam apenas um vazamento de 0,5 mm. Novamente, o segundo vazamento só seria acusado após tratamento do primeiro.

A Rede 6, com dados de m13, o microfone mais distante dos vazamentos, apresentou resultados semelhantes a Rede 5 em áreas distinguidas para identificações de vazamentos.

Assim, todas as redes mostraram ser capazes de distinguir uma situação de vazamento, porém, quanto mais longe o microfone dos vazamentos, com mais limitações se torna a identificação. Em situações de vazamentos simultâneos, se o

microfone se encontra mais próximo do vazamento maior, é possível que haja dificuldades em detectar a ocorrência de um vazamento menor mais longínquo.

A rede Kohonen, pelas redes 2 a 6, se mostrou então apta a identificações de dimensão de vazamentos, porém algumas limitações foram encontradas.

O teste de interpolação realizado, retirando todos os experimentos com F3, vazamento de dimensão de 2mm de diâmetro, do treinamento, com dados de m1, mostrou que os dados de teste de F3 ativaram os neurônios com maiores similaridades: vazamentos simultâneos de 1 mm (F1 e F4) e de vazamentos simultâneos de 1 mm e 0,5 mm (F12 e F17). Na Rede 2, de dados de m1 com dados F3 presentes no treinamento, percebe-se que as áreas de vazamentos com F3 fazem fronteiras com estas mesmas configurações experimentais, reforçando a similaridade nos padrões.

O teste de extrapolação, numa situação prática, se mostrou facilmente identificado por acusar um vazamento de 4 mm de diâmetro com situação sem vazamento com batidas no botijão e na tubulação. O comportamento no tempo de batidas e de vazamentos são diferentes, acusando ser um erro de identificação da rede e que o resultado deve ser ignorado. Em casos de extrapolações, outros dados disponíveis são de extrema importância para o descarte do diagnóstico; fazendo com que estes outros dados (como o comportamento sonoro no domínio do tempo) sejam sempre avaliados para verificar concordância com as respostas da rede.

O teste de dados de m6 em rede treinada por dados do microfone m2, um microfone a uma posição diferente daquele usado no treinamento da rede, mostrou que as situações de não vazamento foram as maiores concordâncias, mostrando que ao se testar dados a uma distância em relação ao microfone diferente daquela dos dados usados no treinamento, ainda é possível haver identificação de padrão para “sem vazamento” para “evento anormal”.

Os testes realizados sem PCA, com frequências de 100 Hz à 1.600 Hz, mostraram que a rede de Kohonen era capaz de identificar áreas para neurônios ativados por respectivas configurações de vazamentos simulados com menos e menores limitações do que quando realizada a compressão dos dados. De fato, houve uma maior facilidade em associar áreas de neurônios na matriz U em relação às situações simuladas do que nas Redes 2 a 6. Porém, a escolha randômica da frequência faz com que a amostragem possa ser não representativa dos dados. Assim, também se faz necessário uma escolha não aleatória das frequências importantes para o reconhecimento do sistema.

As respostas com dados apenas em regime permanente mostraram uma possibilidade de usar menos dados para a identificação de vazamentos não considerando dados no regime transiente, nesse sentido a identificação se limita a não poder caracterizar o vazamento pelo tempo em que o sistema esteja em transiente.

Assim, o estudo com o método acústico se mostrou com muitas informações e muitas alternativas de estudo. No presente trabalho, foi possível identificar o vazamento e caracterizar o sistema apenas pela resposta acústica. A redução da dimensionalidade diminuiu o esforço computacional- para processar os dados- com alta representatividade, mas resultados de testes sem a utilização de PCA sugerem melhores resultados por outros caminhos de redução da matriz de dados para treinamento da rede.

As redes neurais não supervisionadas se mostraram eficientes para o mapeamento dos padrões de cada situação estudada. Ela ainda se mostra como uma alternativa potencialmente melhor do que usar redes neurais supervisionadas para o reconhecimento dos padrões, pois por se tratar de uma rede neural não supervisionada, há o reconhecimento de padrões nos dados, quaisquer que sejam, por mais complexo que se torne o sistema, sem a necessidade de se associar valores numéricos às saídas do sistema.

REFERÊNCIAS

- [1] F. G. Cavalieri, Sistema de Localização de Vazamentos de Gas por Dispositivo Acústico-Eletrônico de Deslocamento Interno [Dissertação de Mestrado], Campinas: Unicamp, 2015.
- [2] A. H. A. Baghdadi e H. A. Mansy, “A mathematical model for leak location in pipelines,” *Applied Mathematical Modeling*, vol. 12, pp. 15-30, 1988.
- [3] S. Liag, “Mathematical modeling of the flow in a pipeline with a leak,” *Mathematics and computers in simulation*, pp. 2253-2267, 2012.
- [4] S. Dudic, I. Ignjatovic, D. Seslija, V. Blagojevic e M. Stojiljkovic, “Leakage quantification of compressed air using ultrasound and infrared thermography,” *Measurement*, vol. 45, pp. 1689-1694, 2012.
- [5] R. B. Santos, Detecção de Vazamentos em tubulações de gás de baixa pressão por meio de sensores acústicos e processamento de sinais [Tese de Doutorado], Campinas: Unicamp, 2015.
- [6] A. M. Pavan, Detecção on-line de vazamentos em vasos de pressão utilizando sistemas sônicos [Dissertação de Mestrado], Campinas: Unicamp, 2005.
- [7] E. O. de Sousa, Detecção de vazamentos em tubulações através do método acústico e da análise dos transientes de pressão [Dissertação de Mestrado], Campinas: Unicamp, 2007.
- [8] J. Kronenburger e J. Sebeson, Analog and Digital Signal Processing: An Integrated Computational Approach With Matlab, Delmar Learning, 2008.
- [9] B. P. Lathi, Sinais e Sistemas Lineares, Bookman, 2004.
- [10] T. G. Fernandes e A. N. Panazio, “Do Analógico ao Digital: Amostragem, Quatização e Codificação,” em *II Simpósio de Iniciação Científica da Universidade Federal do ABC*, Santo André-SP, 2009.
- [11] J. Cooley e J. Tukey, “An Algorithm for the Machine Calculation of Complex Fourier Series,” *Math. Comput.*, vol. 19, nº 90, pp. 297-301, 1965.
- [12] R. G. Lyon, Understanding Digital Signal Processing, Ann Arbor: Prentice Hall, 2011.

- [13] R. Woods, J. McAllister, Y. Yi e G. Lightbody, *FPGA-based Implementation of Signal Processing Systems*, West Sussex: John Wiley & Sons, 2008.
- [14] R. B. Santos, W. S. de Almeida, F. da Silva, S. L. da Cruz e A. M. F. Fileti, "Spectral analysis for detection of leaks in pipes carrying compressed air," *Chemical Engineering Transactions*, vol. 32, pp. 1363-1368, 2013.
- [15] K. Pearson, "On lines and Planes of the closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, nº 11, pp. 559-572, 1901.
- [16] J. E. Jackson, *A user's guide to Principal Components*, Chicago: John Wiley & Sons, Inc, 1991.
- [17] L. H. Chiang, E. L. Russel e R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Londres: Springer-Verlag London Ltd, 2002.
- [18] T. A. Neves, W. S. de Oliveira e J. de Groote, "Sistema de Controle de Acesso Baseado no Reconhecimento de Faces," em *VI Workshop de Visão Computacional*, Presidente Prudente-SP, 2010.
- [19] I. R. T. de Jesus, *Análise de Componentes Principais do Eletromiograma no Domínio da Frequência em Exercícios Cicloergométricos [Dissertação de Mestrado]*, Rio de Janeiro: UFRJ, 2010.
- [20] I. T. Jolliffe, *Principal Components Analysis*, Nova Iorque: Springer-Verlag, 2002.
- [21] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New Jersey: Prentice-Hall International, Inc, 1994.
- [22] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, pp. 338-358, Junho 1965.
- [23] L. A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decisions Process," *Transactions on Systems, Man, and Cybernets*, pp. 28-44, janeiro 1973.
- [24] R. B. Santos, M. Rupp e S. J. Bonzi, "Comparison between multilayer feedforward neural networks and a radial basis function network to detect and locate leaks in pipelines transporting gas," *Chemical Engineering Transactions*, vol. 32, pp. 1375-1380, 2013.
- [25] F. A. Sala, F. M. de Azevedo e F. I. M. Argoud, "Mapas de Kohonen na detecção de eventos epileptogênicos," *IV Workshop de Informática aplicada à Saúde-CBCComp*, pp. 606-609, 2004.

- [26] T. Kohonen, *Self-Organization and Associative Memory*, Berlin: Springer-Verlag, 1989.
- [27] J. Vesanto, J. Himberg, E. Alhoniemi e J. Parhankangas, “Som Toolbox for Matlab 5,” Libella Oy, Espoo, Finlândia, 2000.
- [28] W. Natita, W. Wiboonsak e S. Dusadee, “Appropriate learning rate and neighborhood function of self-organizing map (som) for specific humidity patten classification over southern Thailand,” *International Journal of Modeling and Optimization*, vol. 6, n° 1, pp. 61-65, 2016.
- [29] “Introduction to Self-Organizing Maps,” Material Didático, Universidade de Nova de Lisboa, Instituto Superior de Estatística e Gestão de Informação, [Online]. Available: <http://edugi.uji.es/Bacao/SOM%20Tutorial.pdf>. [Acesso em maio 2016].
- [30] “SOM Toolbox Som_train_struct,” [Online]. Available: http://www.cis.hut.fi/somtoolbox/package/docs2/som_train_struct.html. [Acesso em agosto 2016].
- [31] M. A. Santos da Silva, A. M. Vieira Moteiro e J. S. Medeiros, “Visualization of Geospatial data by Component planes and U-matrix,” *Geoinfo*, Campos do Jordão, 2004.
- [32] “SOM Toolbox som_umat,” [Online]. Available: http://www.cis.hut.fi/somtoolbox/package/docs2/som_umat.html. [Acesso em agosto 2016].
- [33] “SOM Toolbox,” Matlab, [Online]. Available: <https://www.mathworks.com/matlabcentral/linkexchange/links/949>. [Acesso em dezembro 2016].
- [34] “SOM Toolbox Som_make,” [Online]. Available: http://www.cis.hut.fi/somtoolbox/package/docs2/som_make.html. [Acesso em agosto 2016].

APÊNDICE A

A seguinte programação em Matlab (documento .m) possui um passo-a-passo de criação de rede não supervisionada de Kohonen e da criação da matriz U.

A rotina gerou dados relativos a 2 objetos: Objeto 1 e Objeto 2, em relação a 2 atributos: Atributo 1 e Atributo 2. Os dados foram criados randomicamente.

```
%PASSO 1
%Gerando dados
clear;clc;

%Objeto 1; Objeto 2
D1 = rand(15,1);E1 = rand(15,1)+2.5;
D2=rand(15,1)+5;E2= rand(15,1)+2;

D=[D1; D2];%primeiro atributo
E=[E1;E2];%segundo atributo

Dados=[D E];

%%
%PASSO 2
%Iniciando a rede:

msize = [3 3];

%Inicialização linear dos pesos:
sMap = som_lininit(Dados, 'msize', msize, 'lattice', 'hexa');

subplot(1,3,1)

som_grid(sMap, 'Coord', sMap.codebook, 'LineColor', [0.6 0.6 0.5], ...
        'MarkerSize', 4, 'MarkerColor', 'b')

xlabel('Atributo 1');ylabel('Atributo 2');
grid on

subplot(1,3,2)

x=sMap.codebook(:,1);
y=sMap.codebook(:,2);
plot(x,y, '.')
grid on; rotate3d on; hold on

plot(Dados(1:15,1),Dados(1:15,2), '+m');
hold on

plot(Dados(16:30,1),Dados(16:30,2), 'sg');
legend('neur. inicializados','Objeto 1','Objeto 2');

xlabel('Atributo 1');ylabel('Atributo 2');
```

```

%PASSO 3
%TREINAMENTO DA REDE

%treinamento em Batelada dos pesos

sMap = som_batchtrain(sMap,Dados,'trainlen',1000,'radius',[10 1]);
[qe,te] = som_quality(sMap, Dados)%erro de quant. e topografico

%Visualização dos neurônios resultantes do treinamento:

subplot(1,3,3)
x=sMap.codebook(:,1);
y=sMap.codebook(:,2);
plot(x,y, '.')
grid on; rotate3d on; hold on

hold on,
plot(Dados(1:15,1),Dados(1:15,2),'+m');

plot(Dados(16:30,1),Dados(16:30,2),'sg');
legend('neur. treinados','Objeto 1','Objeto 2');

xlabel('Atributo 1');ylabel('Atributo 2');
%%

```

O gráfico resultante desta primeira seção está na Figura A.1.

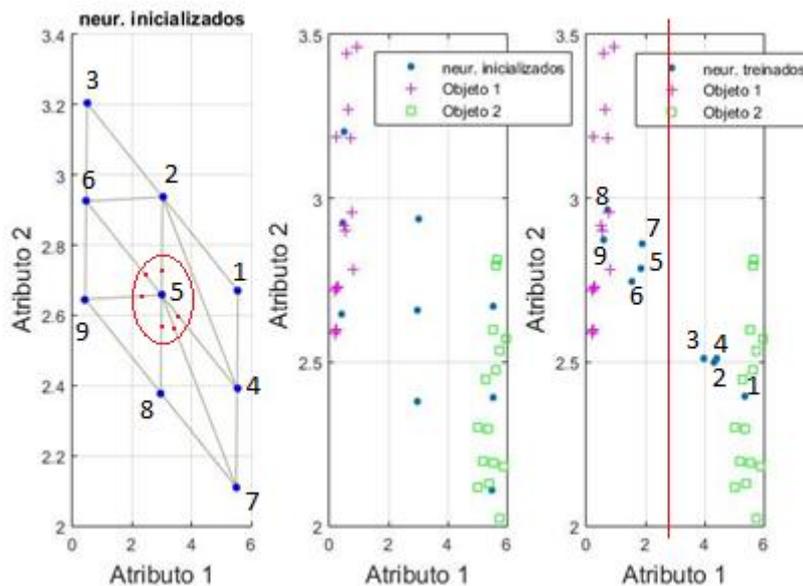


Figura A.1 Dados de treinamento e neurônios

Na figura à esquerda, apresentam-se apenas os neurônios inicializados com a exposição das conexões de um neurônio ao outro. A inicialização escolhida é a linear. A

rede possui 9 neurônios no formato 3x3. A numeração dos neurônios está evidenciada nas figuras da esquerda e da direita, antes e depois do treinamento. Para efeitos de se observar distâncias de um ponto a outro nos gráficos, reparar que os eixos “Atributo 1” e Atributo 2” possuem escalas diferentes.

A organização da vizinhança dos neurônios é hexagonal, portanto, há 6 conexões para cada 1 dos neurônios que são vizinhos imediatos, adjacentes. Assim, cada neurônio (que não esteja nas extremidades) possui 6 conexões. À exemplo, o neurônio 5, circulado em vermelho (o único que não se encontra nos extremos), teve suas conexões evidenciadas pelos pontos vermelhos.

Na figura ao centro, os mesmos neurônios inicializados (sem as conexões à mostra) estão representados junto com os dados de treinamento. Enquanto que, à direita (sem conexões à mostra), os neurônios após treinamento estão juntos com os dados de treinamento.

A reta em vermelho acrescentada na figura à direita foi feita para evidenciar que, após treinamento, os neurônios tendem a se associar aos padrões presentes nos dados de treinamento: alguns neurônios se aproximaram do padrão da direita (Objeto 2) da reta e outros, da esquerda (objeto 1).

Na Tabela A.1, encontram-se as coordenadas dos neurônios já treinados e a numeração de 1 a 9 respectiva de cada um. Os valores das coordenadas dos neurônios são utilizados no cálculo da matriz U.

Tabela A.1 Coordenadas dos neurônios treinados

| Neurônio | Valor em ‘Atributo 1’, x | Valor em ‘Atributo 2’, y |
|----------|--------------------------|--------------------------|
| 1 | 5,3681 | 2,3945 |
| 2 | 4,310 | 2,5000 |
| 3 | 3,9934 | 2,5101 |
| 4 | 4,4209 | 2,5118 |
| 5 | 1,8270 | 2,7867 |
| 6 | 1,5617 | 2,7443 |
| 7 | 1,8831 | 2,8629 |
| 8 | 0,7140 | 2,9645 |
| 9 | 0,5970 | 2,8732 |

A continuação a seguir do programa é a visualização da matriz U.

```

%%
%PASSO 4

%Visualização da matriz U

figure()
colormap(1-gray)
%som_show(sMap,'norm','d')
som_show(sMap,'umat','all')

%matriz com valores calculados da matriz U:

umatrix=som_umat(sMap);

%%

```

A figura da matriz U resultante está representada abaixo.

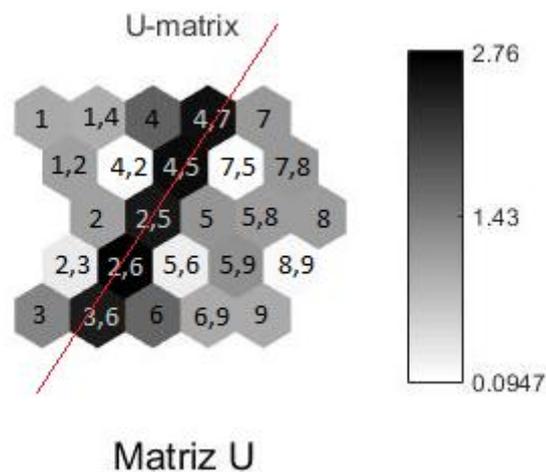


Figura A.2 Matriz U

Na Figura A.2, foram acrescentadas as numerações: de “1” a “9” aos hexágonos representando o respectivo neurônio de 1 a 9; aos hexágonos representando distâncias euclidianas de um neurônio a outro, foram escritas as numerações dos dois neurônios em questão.

A coloração de cada hexágono procede como a seguir:

1. A coloração dos hexágonos de 1 a 9: calcula-se a distância euclidiana do neurônio em questão a todos os neurônios vizinhos imediatos. Em seguida, o valor da mediana destas distâncias corresponde a uma cor (barra ao lado da matriz). Colore-se com a cor correspondente.

Exemplo: o neurônio 1 possui o neurônio 2 e o 4 como vizinhos imediatos, também de acordo com as conexões dos neurônios vistas na Figura A.1. A distância euclidiana, após treinamento, do neurônio 1 ao 2, $d_{1,2}$, é de 1,0635 e $d_{1,4}$, de 1,0088. A mediana destes dois números é o valor 1,0088. Então, procura-se a cor equivalente a 1,0088 e com esta se colore o hexágono de numeração “1”.

A matriz U também pode ser calculada de acordo com a média, com o valor máximo ou mínimo e, como então, com a mediana.

2. As colorações dos hexágonos representando distâncias entre dois neurônios: calcula-se a distância euclidiana entre os 2 neurônios em questão. Diretamente com este valor, verifica-se a cor corresponde e se colore o hexágono.

Exemplo: a distância euclidiana $d_{1,2}$ é de 1,0635. Colore-se o hexágono “1,2” com a coloração correspondente de acordo com a escala de cores ao lado da matriz U.

Na Figura A.2, os hexágonos de coloração preta indicam um distanciamento entre os neurônios 4 e 7, 4 e 5, 2 e 5, 2 e 6 e 3 e 6, a reta vermelha foi acrescentada para evidenciar a separação. Por outro lado, os outros hexágonos possuem coloração mais clara, indicam proximidade. Por exemplo, verifica-se que o neurônio 2 e 4 são bastante próximos pela coloração branca do hexágono “4,2”. Em comparação com a Figura A.1, as duas retas vermelhas acabam por separar mesmos grupos: de um lado “1”, “2”, “3” e “4” e, do outro, “5”, “6”, “7”, “8” e “9”.

A vantagem de analisar os dados em matriz U é que pode ser sempre representada em poucas dimensões, como 1-D e 2-D. A análise da coloração total da matriz U sugere o agrupamento dos neurônios em duas regiões distintas, pois há uma coluna de hexágonos escuros separando duas regiões mais claras (neurônios próximos).

A parte final da programação evidencia os neurônios ativado pelos dados de treinamento da rede. O neurônio que melhor representa cada dado de treino é ativado, ou seja, o neurônio mais próximo, por distância Euclidiana, a cada dado de entrada.

```
%%
%PASSO 5
%Evidenciar ativação pelos dados de treino
Obj1=[D1 E1];
h1 = som_hits(sMap,Obj1);%dados de "Objeto 1"

Obj2=[D2 E2];
h2 = som_hits(sMap,Obj2);%dados de "Objeto 2"

som_show_add('hit', [h1,h2], 'MarkerColor', ...
    [1 0 1; 0 1 0], 'Subplot',1)
%1-rosa escuro, 2-verde
```

%FIM

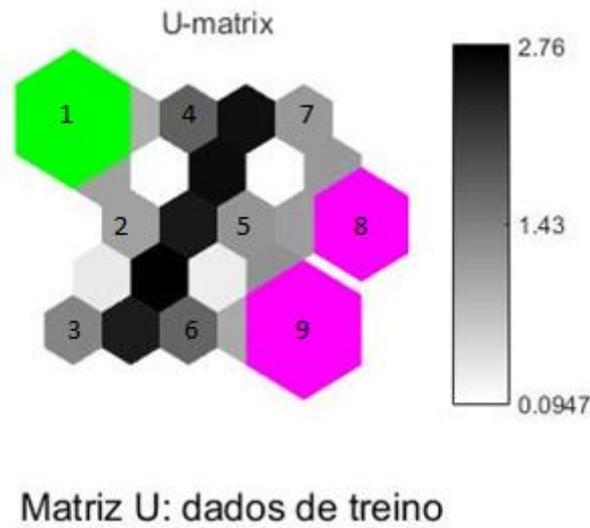


Figura A.3 Neurônios ativados pelos dados de treino. Verde: dados de Objeto 2; Rosa: dados de Objeto 1

A Figura A.3 mostra que todos os dados relativos ao Objeto 2 ativam o neurônio “1”, em verde, e que os dados do Objeto 1 ativam os neurônios 8 e 9, em rosa. Quanto mais vezes forem ativados, maiores se tornam as representações dos hexágonos. Assim, nesta representação, dentre os neurônios ativados, o neurônio 8 foi menos vezes ativado.

APÊNDICE B

A programação em Matlab (documento .m) para treinamento e testes da Rede 4, de m6.

```

%Criação da rede de Kohonen com m6. Após PCA.
clear;
clc;
lat=load('latent.dat'); %carrega autovalores (do PCA)
pcs=load('score500.dat');%carrega dados em termos de C.P.
pcs=pcs(:,1:200);%200 cp usadas
repre=sum(lat(1:200))/sum(lat);%200 pcs

%%
%Grafico da representatividade
n=500;
for i=1:n
rep=(sum(lat(1:i))/sum(lat))*100;
plot(i,rep,'k.')
hold on
i=i+1;
end
grid on
title ('Representatividade dos dados em Componentes Principais %');
xlabel('Número de Componentes'); ylabel('Percentual Cumulativo de
Variância');
%%

%normalizar
pcs = som_normalize(pcs,'histC');%histC

%Pegar dados de m6 (terceiro microf. usado)
m3_sv=pcs(147:219,:);
m3_sv_bb=pcs(512:584,:);
m3_sv_bt=pcs(877:949,:);
m3_F1=pcs(1242:1314,:);
m3_F2=pcs(1607:1679,:);
m3_F3=pcs(1972:2044,:);
m3_F4=pcs(2337:2409,:);
m3_F5=pcs(2702:2774,:);
m3_F6=pcs(3067:3139,:);
m3_F7=pcs(3432:3504,:);
m3_F1_F2=pcs(3797:3869,:);
m3_F1_F3=pcs(4162:4234,:);
m3_F1_F4=pcs(4527:4599,:);
m3_F1_F7=pcs(4892:4964,:);
m3_F2_F7=pcs(5257:5329,:);

%Dados de treino m3
m3_sv_tr=m3_sv(1:50,:);
m3_sv_bb_tr=m3_sv_bb(1:50,:);
m3_sv_bt_tr=m3_sv_bt(1:50,:);
m3_F1_tr=m3_F1(1:50,:);
m3_F2_tr=m3_F2(1:50,:);
m3_F3_tr=m3_F3(1:50,:);

%Dados de teste m3
m3_sv_tt=m3_sv(51:73,:);
m3_sv_bb_tt=m3_sv_bb(51:73,:);
m3_sv_bt_tt=m3_sv_bt(51:73,:);
m3_F1_tt=m3_F1(51:73,:);
m3_F2_tt=m3_F2(51:73,:);
m3_F3_tt=m3_F3(51:73,:);

```

```

m3_F4_tr=m3_F4(1:50,:);
m3_F5_tr=m3_F5(1:50,:);
m3_F6_tr=m3_F6(1:50,:);
m3_F7_tr=m3_F7(1:50,:);
m3_F1_F2_tr=m3_F1_F2(1:50,:);
m3_F1_F3_tr=m3_F1_F3(1:50,:);
m3_F1_F4_tr=m3_F1_F4(1:50,:);
m3_F1_F7_tr=m3_F1_F7(1:50,:);
m3_F2_F7_tr=m3_F2_F7(1:50,:);

m3_F4_tt=m3_F4(51:73,:);
m3_F5_tt=m3_F5(51:73,:);
m3_F6_tt=m3_F6(51:73,:);
m3_F7_tt=m3_F7(51:73,:);
m3_F1_F2_tt=m3_F1_F2(51:73,:);
m3_F1_F3_tt=m3_F1_F3(51:73,:);
m3_F1_F4_tt=m3_F1_F4(51:73,:);
m3_F1_F7_tt=m3_F1_F7(51:73,:);
m3_F2_F7_tt=m3_F2_F7(51:73,:);

```

```

m3_tr=[m3_sv_tr;m3_sv_bb_tr;m3_sv_bt_tr;m3_F1_tr;m3_F2_tr;m3_F3_tr;...
m3_F4_tr; m3_F5_tr;m3_F6_tr;m3_F7_tr;m3_F1_F2_tr;m3_F1_F3_tr;...
m3_F1_F4_tr;m3_F1_F7_tr;m3_F2_F7_tr];

```

```

m3_tt=[m3_sv_tt;m3_sv_bb_tt;m3_sv_bt_tt;m3_F1_tt;m3_F2_tt;m3_F3_tt;...
m3_F4_tt; m3_F5_tt;m3_F6_tt;m3_F7_tt;m3_F1_F2_tt;m3_F1_F3_tt;...
m3_F1_F4_tt;m3_F1_F7_tt;m3_F2_F7_tt];

```

```

sDpcs = som_data_struct(m3_tr,'name','data in pc',...
'comp_names',{'p.c.1','p.c.2','p.c.3','p.c.4','p.c.5',...
'p.c.6','p.c.7','p.c.8','p.c.9','p.c.10','p.c.11',...
'p.c.12','p.c.13','p.c.14','p.c.15','p.c.16','p.c.17',...
'p.c.18','p.c.19','p.c.20','p.c.21','p.c.22','p.c.23',...
'p.c.24','p.c.25','p.c.26','p.c.27','p.c.28','p.c.29',...
'p.c.30','p.c.31','p.c.32','p.c.33','p.c.34','p.c.35',...
'p.c.36','p.c.37','p.c.38','p.c.39','p.c.40',...
'p.c.41','p.c.42','p.c.43','p.c.44','p.c.45','p.c.46',...
'p.c.47','p.c.48','p.c.49','p.c.50','p.c.51','p.c.52',...
'p.c.53','p.c.54','p.c.55','p.c.56','p.c.57','p.c.58',...
'p.c.59','p.c.60','p.c.61','p.c.62','p.c.63','p.c.64',...
'p.c.65','p.c.66','p.c.67','p.c.68','p.c.69','p.c.70',...
'p.c.71','p.c.72','p.c.73','p.c.74','p.c.75','p.c.76',...
'p.c.77','p.c.78','p.c.79','p.c.80','p.c.81','p.c.82',...
'p.c.83','p.c.84','p.c.85','p.c.86','p.c.87','p.c.88',...
'p.c.89','p.c.90','p.c.91','p.c.92','p.c.93','p.c.94',...
'p.c.95','p.c.96','p.c.97','p.c.98','p.c.99','p.c.100',...
'p.c.101','p.c.102','p.c.103','p.c.104','p.c.105',...
'p.c.106','p.c.107','p.c.108','p.c.109','p.c.110','p.c.111',...
'p.c.112','p.c.113','p.c.114','p.c.115','p.c.116','p.c.117',...
'p.c.118','p.c.119','p.c.120','p.c.121','p.c.122','p.c.123',...

```

```

'p.c.124','p.c.125','p.c.126','p.c.127','p.c.128','p.c.129',...
'p.c.130','p.c.131','p.c.132','p.c.133','p.c.134','p.c.135',...
    'p.c.136','p.c.137','p.c.138','p.c.139','p.c.140',...
'p.c.141','p.c.142','p.c.143','p.c.144','p.c.145','p.c.146',...
'p.c.147','p.c.148','p.c.149','p.c.150','p.c.151','p.c.152',...
'p.c.153','p.c.154','p.c.155','p.c.156','p.c.157','p.c.158',...
'p.c.159','p.c.160','p.c.161','p.c.162','p.c.163','p.c.164',...
'p.c.165','p.c.166','p.c.167','p.c.168','p.c.169','p.c.170',...
'p.c.171','p.c.172','p.c.173','p.c.174','p.c.175','p.c.176',...
'p.c.177','p.c.178','p.c.179','p.c.180','p.c.181','p.c.182',...
'p.c.183','p.c.184','p.c.185','p.c.186','p.c.187','p.c.188',...
'p.c.189','p.c.190','p.c.191','p.c.192','p.c.193','p.c.194',...
'p.c.195','p.c.196','p.c.197','p.c.198','p.c.199','p.c.200'});
    sDpcs = som_label(sDpcs,'add',[1:50]','SV'); %sem vazamento
    sDpcs = som_label(sDpcs,'add',[51:100]','SVB');%sem vaz com bat
botijao
    sDpcs = som_label(sDpcs,'add',[101:150]','SVT');%sem vaz com bat tub
    sDpcs = som_label(sDpcs,'add',[151:200]','F1');%F1_1mm
    sDpcs = som_label(sDpcs,'add',[201:250]','F2');%F2_0,5mm
    sDpcs = som_label(sDpcs,'add',[251:300]','F3');%F3_2mm
    sDpcs = som_label(sDpcs,'add',[301:350]','F4');%F4_1mm
    sDpcs = som_label(sDpcs,'add',[351:400]','F5');%F5_4mm
    sDpcs = som_label(sDpcs,'add',[401:450]','F6');%F6_3mm
    sDpcs = som_label(sDpcs,'add',[451:500]','F7');%F7_0.5mm
    sDpcs = som_label(sDpcs,'add',[501:550]','F12');%F1_F2
    sDpcs = som_label(sDpcs,'add',[551:600]','F13');%F1_F3
    sDpcs = som_label(sDpcs,'add',[601:650]','F14');%F1_F4
    sDpcs = som_label(sDpcs,'add',[651:700]','F17');%F1_F7
    sDpcs = som_label(sDpcs,'add',[701:750]','F27');%F2_F7

%%
%criacao da rede
sMap = som_make(sDpcs); %configuração de base, pode pular esta secao

%%

msize=[22, 10];
sMap = som_lininit(sDpcs, 'msize', msize,'shape','sheet');

sMap
=
som_batchtrain(sMap,sDpcs,'trainlen',3000,'neigh','cutgauss',...
    'shape','sheet');

sMap = som_autolabel(sMap,sDpcs,'vote'); %coloca legenda na matriz U

[qe,te] = som_quality(sMap,sDpcs) %qualidade da rede (E.Q e E.T.)

%%

```

```

%neuronios treinados (3 primeiras dimensoes)
simb={'b+', 'y.', 'gs', 'rp', 'kd', 'o', '>m', 'y*', 'm.', 'cx'};
x=sMap.codebook(:,1);
y=sMap.codebook(:,2);
z=sMap.codebook(:,3);
plot3(x,y,z, '.')
grid on; rotate3d on; hold on
%dados de entrada:
%sem vazamento
x=sDpcs.data(1:50,1);y=sDpcs.data(1:50,2);z=sDpcs.data(1:50,3);
plot3(x,y,z,simb{1});hold on
%sem vazamneto bat botijao

x=sDpcs.data(51:100,1);y=sDpcs.data(51:100,2);z=sDpcs.data(51:100,3);
plot3(x,y,z,simb{2}); hold on
%sem vazamneto bat tubulacao

x=sDpcs.data(101:150,1);y=sDpcs.data(101:150,2);z=sDpcs.data(101:150,3);
plot3(x,y,z,simb{3}); hold on
%F1_1mm
x=sDpcs.data(151:200,1);y=sDpcs.data(151:200,2);
z=sDpcs.data(151:200,3);
plot3(x,y,z,simb{4}); hold on
%F2_0.5mm
x=sDpcs.data(201:250,1);y=sDpcs.data(201:250,2);
z=sDpcs.data(201:250,3);
plot3(x,y,z,simb{5}); hold on
%F3_2mm
x=sDpcs.data(251:300,1);y=sDpcs.data(251:300,2);
z=sDpcs.data(251:300,3);
plot3(x,y,z,simb{6}); hold on
%F4_1mm
x=sDpcs.data(301:350,1);y=sDpcs.data(301:350,2);
z=sDpcs.data(301:350,3);
plot3(x,y,z,simb{7}); hold on
%F5_4mm
x=sDpcs.data(351:400,1);y=sDpcs.data(351:400,2);
z=sDpcs.data(351:400,3);
plot3(x,y,z, '*m'); hold on
%F6_3mm
x=sDpcs.data(401:450,1);y=sDpcs.data(401:450,2);
z=sDpcs.data(401:450,3);
plot3(x,y,z,simb{8}); hold on
%F7_0.5mm
x=sDpcs.data(451:500,1);y=sDpcs.data(451:500,2);
z=sDpcs.data(451:500,3);
plot3(x,y,z,simb{10}); hold on
%F1_F2
x=sDpcs.data(501:550,1);y=sDpcs.data(501:550,2);
z=sDpcs.data(501:550,3);
plot3(x,y,z,simb{1}); hold on
%F1_F3
x=sDpcs.data(551:600,1);y=sDpcs.data(551:600,2);
z=sDpcs.data(551:600,3);
plot3(x,y,z,simb{2}); hold on
%F1_F4
x=sDpcs.data(601:650,1);y=sDpcs.data(601:650,2);
z=sDpcs.data(601:650,3);
plot3(x,y,z,simb{3}); hold on
%F1_F7

```

```

x=sDpcs.data(651:700,1);y=sDpcs.data(651:700,2);
z=sDpcs.data(651:700,3);
plot3(x,y,z,simb{4}); hold on
%F2_F7
x=sDpcs.data(701:750,1);y=sDpcs.data(701:750,2);
z=sDpcs.data(701:750,3);
plot3(x,y,z,simb{5}); hold on
xlabel('C.P.1'); ylabel('C.P.2');zlabel('C.P.3');
title('Dados (de ml) usados no treino e neurônios');
legend('Neurônios treinados','Sem vaz.','Sem vaz.; bat. no
botijão','Sem vaz.; bat. na tubulação',...
'F1 1mm','F2 0.5mm','F3 2mm','F4 1mm','F5 4mm',...
'F6 3mm','F7 0.5mm','F1 F2','F1 F3','F1 F4','F1 F7','F2 F7')

%%%%%%
%%
%visualizar u-matrix dados de treino, todos as configurações
figure()

colormap(1-gray)
subplot(1,2,1)
%som_show(sMap,'norm','d')
som_show(sMap,'umat','all')

%dados treinados de 1 furo apenas (e sem vazamento)

h1 = som_hits(sMap,sDpcs.data(1:50,:));%%sem vazamento
h2 = som_hits(sMap,sDpcs.data(51:100,:));%sem vazamento_bat_botijao
h3 = som_hits(sMap,sDpcs.data(101:150,:));%sem vazamento bat tub
h4 = som_hits(sMap,sDpcs.data(151:200,:));%F1_1mm
h5 = som_hits(sMap,sDpcs.data(201:250,:));%F2_05mm
h6 = som_hits(sMap,sDpcs.data(251:300,:));%F3_2mm
h7 = som_hits(sMap,sDpcs.data(301:350,:));%F4_1mm
h8 = som_hits(sMap,sDpcs.data(351:400,:));%F5_4mm
h9 = som_hits(sMap,sDpcs.data(401:450,:));%F6_3mm
h10 = som_hits(sMap,sDpcs.data(451:500,:));%F7_05mm

%dados de treino de 2 furos simultaneos

h11 = som_hits(sMap,m3_F1_F2_tr);%F1_F2
h12 = som_hits(sMap,m3_F1_F3_tr);%F1_F3
h13 = som_hits(sMap,m3_F1_F4_tr);%F1_F4
h14 = som_hits(sMap,m3_F1_F7_tr);%F1_F7
h15 = som_hits(sMap,m3_F2_F7_tr);%F2_F7

```

```

som_show_add('hit', [h1,h2,h3,h4,
h5,h6,h7,h8,h9,h10,h11,h12,h13,h14,h15], 'MarkerColor',...
    [ 0 0 0;1 0 1; 0 0 0.5 ; 1 0 0; 0 1 0;0 0 1;1 0.5 0;...
    1 1 0;0 1 1; 0.7 0.7 0.7;1 1 0.5;0.3 0.4 0;0.5 0 0.5;...
    1 0.8 0.8; 0.3 0.5 1], 'Subplot',1)
%1-preto, 2-rosa escuro, 3-azul marinho, 4,5,6-RGB, 7-laranja, 8-
amarelo,
%9-ciano (azul-verde claro), 10-cinza,
%11-amarelo claro,12-verde escuro,13-roxo, 14-rosa claro,15-azul
celeste
%%

figure()
som_show(sMap,'umat','all')
colormap(1-gray)

som_show_add('label',sMap,'subplot',1) %Matriz U com legendas

%%
%visualizar u-matrix dados de teste, todos as configurações

figure()
colormap(1-gray)
%som_show(sMap,'norm','d')
som_show(sMap,'umat','all')

%dados treinados de 1 furo apenas (e sem vazamento)

h1 = som_hits(sMap,m3_sv_tt);%%sem vazamento

h2 = som_hits(sMap,m3_sv_bb_tt);%sem vazamento_bat_botijao

h3 = som_hits(sMap,m3_sv_bt_tt);%sem vazamento bat tub

h4 = som_hits(sMap,m3_F1_tt);%F1_1mm

h5 = som_hits(sMap,m3_F2_tt);%F2_05mm

h6 = som_hits(sMap,m3_F3_tt);%F3_2mm

h7 = som_hits(sMap,m3_F4_tt);%F4_1mm

h8 = som_hits(sMap,m3_F5_tt);%F5_4mm

h9 = som_hits(sMap,m3_F6_tt);%F6_3mm

h10 = som_hits(sMap,m3_F7_tt);%F7_05mm

%dados de treino de 2 furos simultaneos

h11 = som_hits(sMap,m3_F1_F2_tt);%F1_F2

h12 = som_hits(sMap,m3_F1_F3_tt);%F1_F3

```

```

h13 = som_hits(sMap,m3_F1_F4_tt);%F1_F4

h14 = som_hits(sMap,m3_F1_F7_tt);%F1_F7

h15 = som_hits(sMap,m3_F2_F7_tt);%F2_F7

som_show_add('hit',[h1,h2,h3,h4,
h5,h6,h7,h8,h9,h10,h11,h12,h13,h14,h15], 'MarkerColor',...
[ 0 0 0;1 0 1; 0 0 0.5 ; 1 0 0; 0 1 0;0 0 1;1 0.5 0;...
1 1 0;0 1 1; 0.7 0.7 0.7;1 1 0.5;0.3 0.4 0;0.5 0 0.5;...
1 0.8 0.8; 0.3 0.5 1], 'Subplot',1)
%1-preto, 2-rosa escuro, 3-azul marinho, 4,5,6-RGB, 7-laranja, 8-
amarelo,
%9-ciano (azul-verde claro), 10-cinza,
%11-amarelo claro,12-verde escuro,13-roxo, 14-rosa claro,15-azul
celeste

%%
%Teste de extrapolação:
%pegar dados apenas do m2 para teste da rede (treinada com dados de
m6)

pcs1_m2=pcs(74:146,:);%sem vazamento
pcs2_m2=pcs(439:511,:);%sem vazamento_bat_botijao
pcs3_m2=pcs(804:876,:);%sem vazamento bat tub
pcs4_m2=pcs(1169:1241,:);%F1_1mm
pcs5_m2=pcs(1534:1606,:);%F2_05mm
pcs6_m2=pcs(1899:1971,:);%F3_2mm
pcs7_m2=pcs(2264:2336,:);%F4_1mm
pcs8_m2=pcs(2629:2701,:);%F5_4mm
pcs9_m2=pcs(2994:3066,:);%F6_3mm
pcs10_m2=pcs(3359:3431,:);%F7_05mm
pcs11_m2=pcs(3724:3796,:);%F1_F2
pcs12_m2=pcs(4089:4161,:);%F1_F3
pcs13_m2=pcs(4454:4526,:);%F1_F4
pcs14_m2=pcs(4819:4891,:);%F1_F7
pcs15_m2=pcs(5184:5256,:);%F2_F7
%%
%visualizar u-matrix
figure()
colormap(1-gray)
%som_show(sMap,'norm','d')
som_show(sMap,'umat','all')

%teste da rede treinada por m6 com dados de outro microfone (m2)
h1 = som_hits(sMap,pcs1_m2);%%sem vazamento

h2 = som_hits(sMap,pcs2_m2);%sem vazamento_bat_botijao

h3 = som_hits(sMap,pcs3_m2);%sem vazamento bat tub

h4 = som_hits(sMap,pcs4_m2);%F1_1mm

h5 = som_hits(sMap,pcs5_m2);%F2_05mm

h6 = som_hits(sMap,pcs6_m2);%F3_2mm

h7 = som_hits(sMap,pcs7_m2);%F4_1mm

```

```
h8 = som_hits(sMap,pcs8_m2);%F5_4mm

h9 = som_hits(sMap,pcs9_m2);%F6_3mm

h10 = som_hits(sMap,pcs10_m2);%F7_05mm

som_show_add('hit',[h1,h2,h3,h4, h5,h6,h7,h8,h9,h10], 'MarkerColor',[ 0
0 0;1 0 1; 0 0 0.5 ; 1 0 0; 0 1 0;0 0 1;1 0.5 0.2;1 1 0;0 1 1; 0.7 0.7
0.7], 'Subplot',1)
%1-preto, 2-rosa, 3-azul marinho, 4,5,6-RGB, 7-laranja, 8-amarelo, 9-
ciano (azul-verde claro), 10-cinza%

%FIM
```

APÊNDICE C

Respostas de configurações experimentais no tempo.

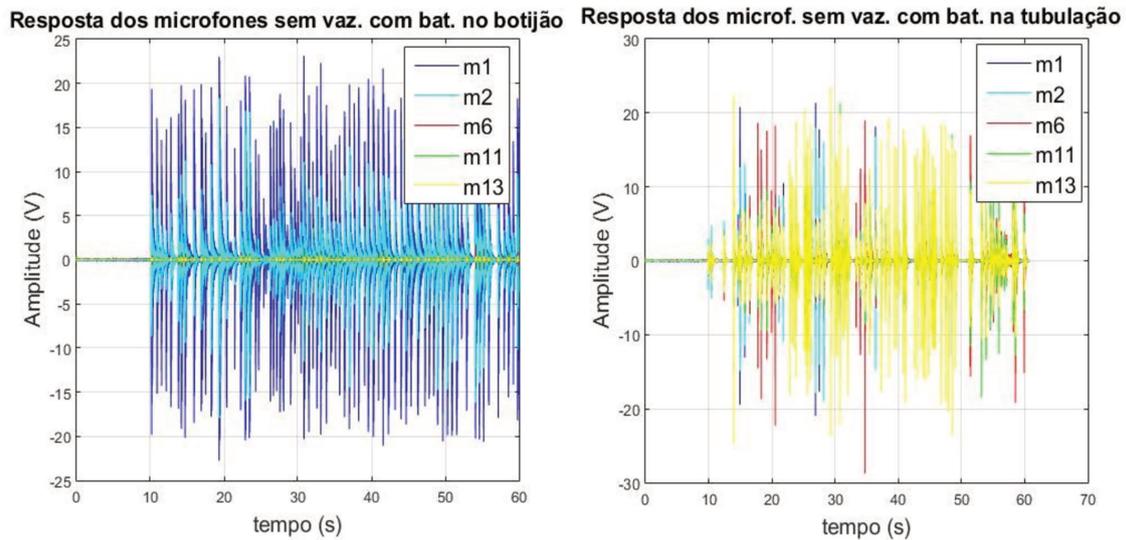


Figura C.1 Batidas no botijão

Figura C.2 Batidas na tubulação

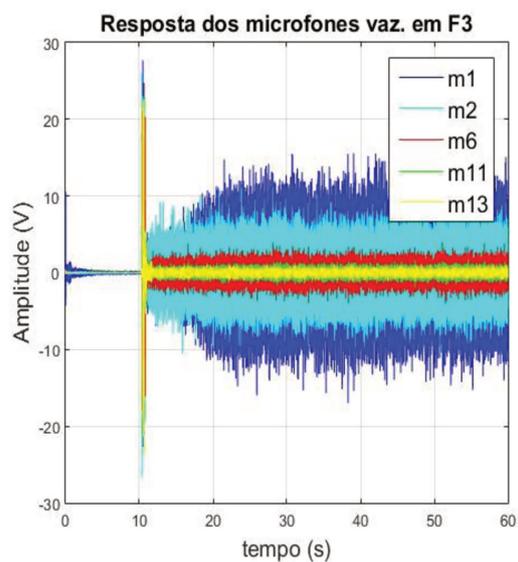


Figura C.3 Vazamento em F3, de 2 mm de diâmetro

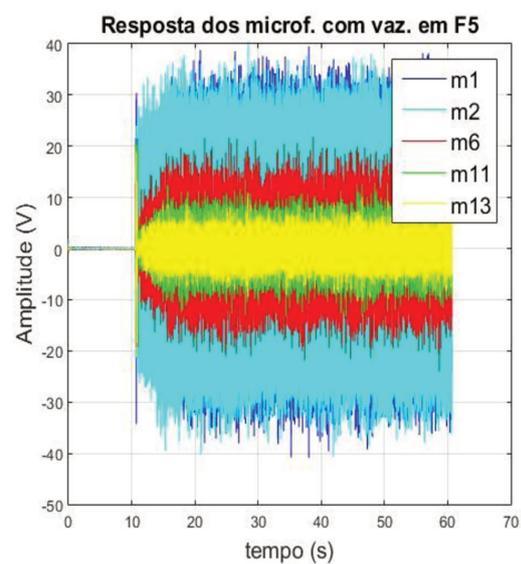


Figura C.4 Vazamento em F5, de 4 mm de diâmetro

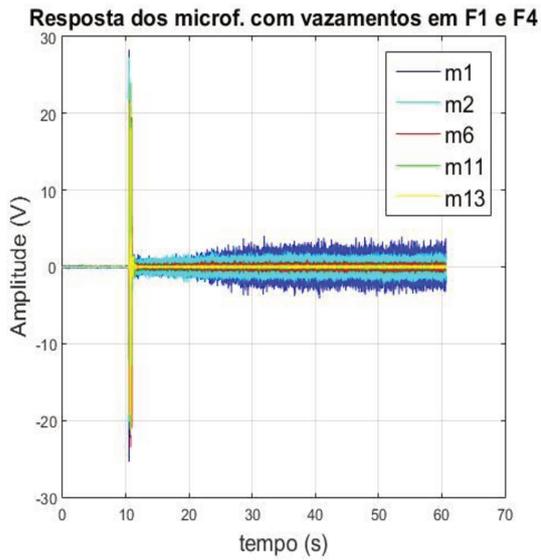


Figura C.5 Vazamentos simultâneos em F1 e em F4, 1 mm de diâmetro cada

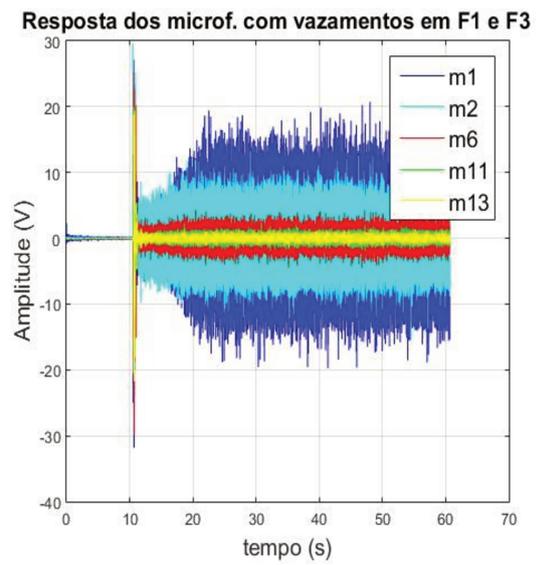


Figura C.6 Vazamentos simultâneos em F1 e em F3, 1 mm e 2 mm de diâmetro

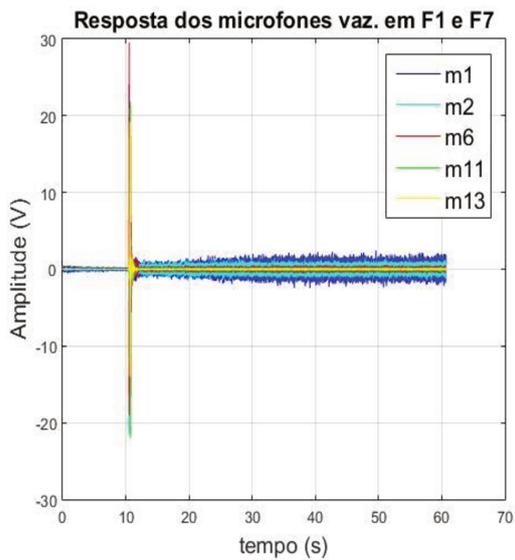


Figura C.7 Vazamentos simultâneos em F1 e em F7, 1 mm e 0,5 mm de diâmetro

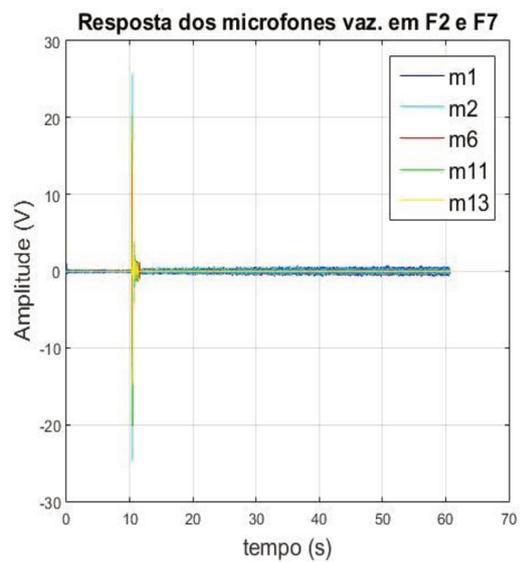


Figura C.8 Vazamentos em F2 e F7 simultâneos, 0,5 mm de diâmetro cada

APÊNDICE D

As respostas no domínio da frequência de vazamento em F5 (4 mm de diâmetro) e de vazamentos simultâneos de F1 e F3 apresentam-se neste apêndice.

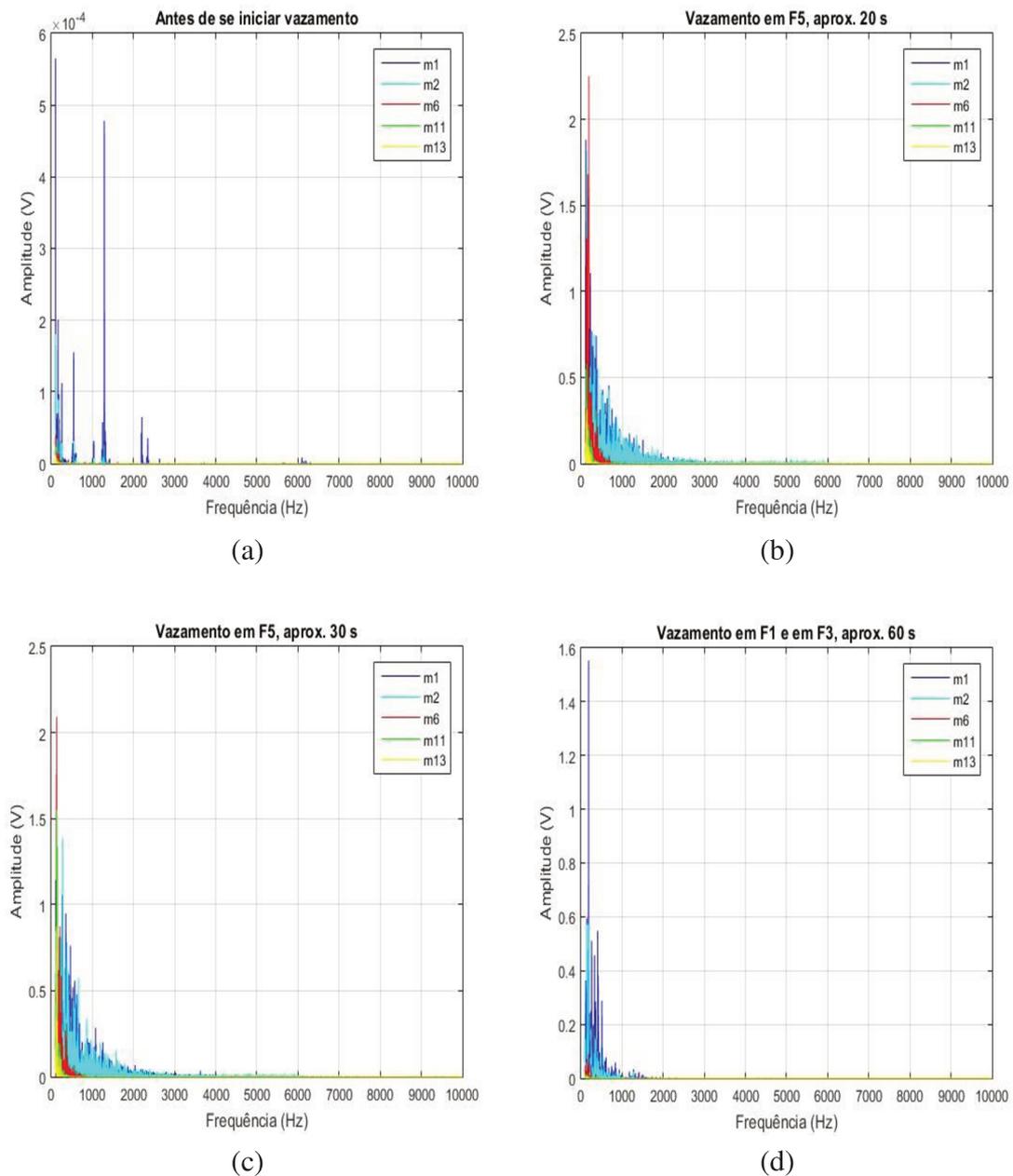


Figura D.1 Vazamento em F4 (4 mm de diâmetro) em tempos (aproximados) diferentes.

(a) antes de simular vazamento; (b) em aproximadamente 20 s; (c) em aproximadamente em 30 s; (d) em aproximadamente em 60 s

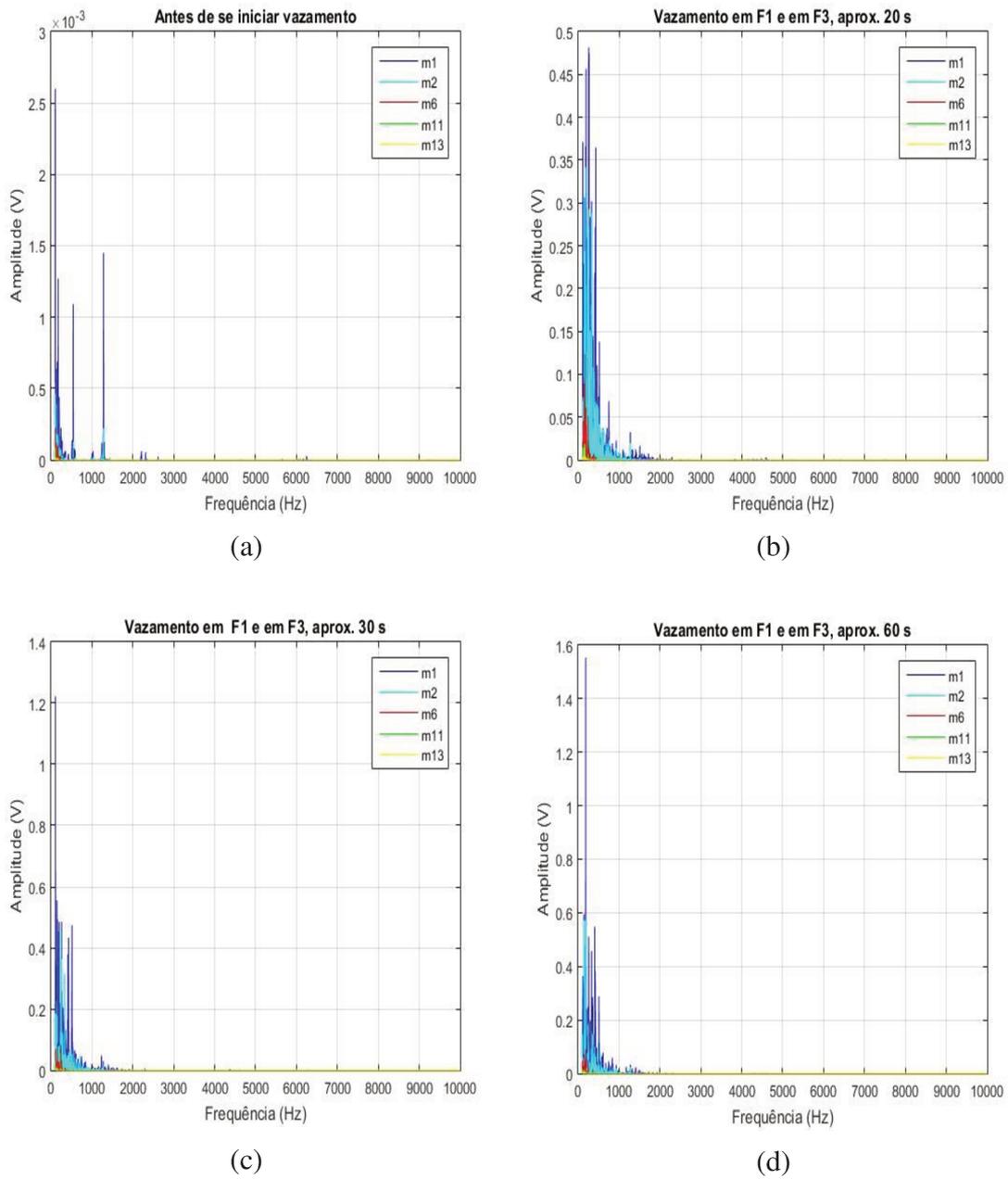


Figura D.2 Vazamento em F1 (1 mm) e em F3 (2 mm) simultâneos em tempos (aproximados) diferentes. (a) antes de simular vazamento; (b) em aproximadamente 20 s; (c) em aproximadamente em 30 s; (d) em aproximadamente em 60 s

APÊNDICE E

Este apêndice mostra outros resultados de extrapolações da rede de Kohonen em relação à distância, ou seja, teste de mapa com dados de um microfone diferente daquele que forneceu os dados de treino.

1) Teste com Dados Comprimidos por PCA

A Figura E.1, com a figura (a) sendo repetida da Figura 4.16 (b) deste trabalho, representa um teste de extrapolação por distância com rede treinada por dados em termos de suas 200 primeiras componentes principais: os dados de teste são provenientes de m13 enquanto que os dados de treino foram os dados de m2 (Rede 3).

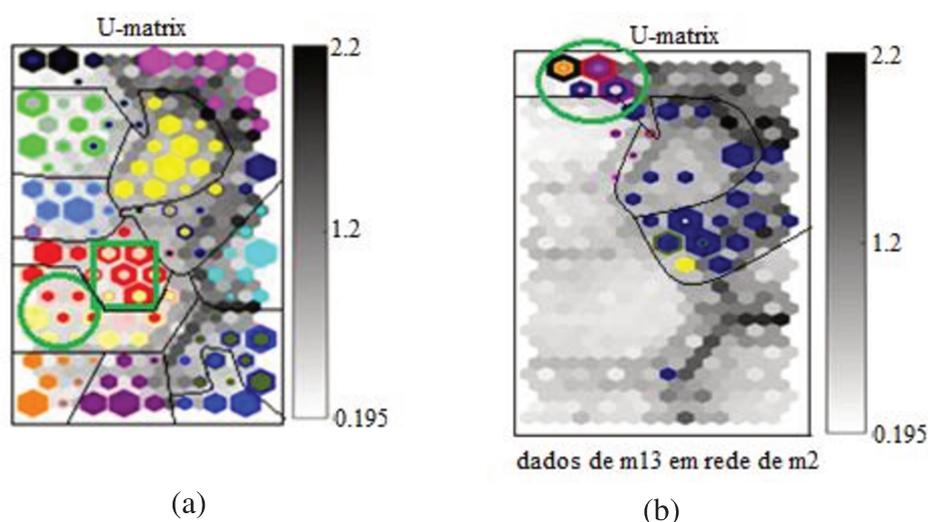


Figura E.1 Representação da matriz U da Rede 3 com (a) neurônios ativados pelos dados de treinamento (figura repetida); (b) neurônios ativados no teste pelos dados do m13, extrapolação

A representação das cores é a mesma dos gráficos anteriores das matrizes U. Na figura (b), fez-se a divisão das áreas seguindo aquela da figura (a). Percebe-se que no teste todos os tipos de vazamentos, em sua maioria no interior do círculo verde em (b), capturados pelo microfone 13 foram considerados como sem vazamento pela rede treinada por dados de m2 (estão na área equivalente a hexágonos em preto, rosa escuro e azul escuro da figura (a)). Mais especificamente, elas coincidem com os hexágonos em preto (sem qualquer anormalidade).

Este resultado provavelmente é causado pois amplitudes acústicas com poucas variações são perceptíveis pelo microfone mais distante em situações de vazamento, m13. No entanto, para o microfone que se encontra perto, como o m2, as variações nas respostas são mais expressivas. Assim, a rede treinada por dados de m2, ao comparar os dados de teste de m13, as amplitudes se mostram mais próximos de situações com ausência de qualquer perturbação.

As situações de batidas na tubulação são as que mais coincidem. Porém, alguns hexágonos que foram associados ao padrão de vazamento em F5 também são ativados por batidas. Nesta situação, o erro causado pela extrapolação poderia ser denunciado ao se avaliar em conjunto o comportamento das amplitudes no tempo, por exemplo, como mencionado anteriormente.

2) Testes com Dados Diretamente do Domínio da Frequência (100 Hz-1600 Hz)

Dos mesmos mapas expostos em Figura 4.29 e a Figura 4.30 com dados de treino e de teste oriundos respectivamente de m1 e de m13, fizeram-se testes com extrapolações de distância.

Dados de Teste de m13 na Rede de Treino com Dados de m1

A Figura E.2 representa a ativação dos neurônios com os dados extrapolados, dessa vez, os dados não passaram por uma compressão (PCA).

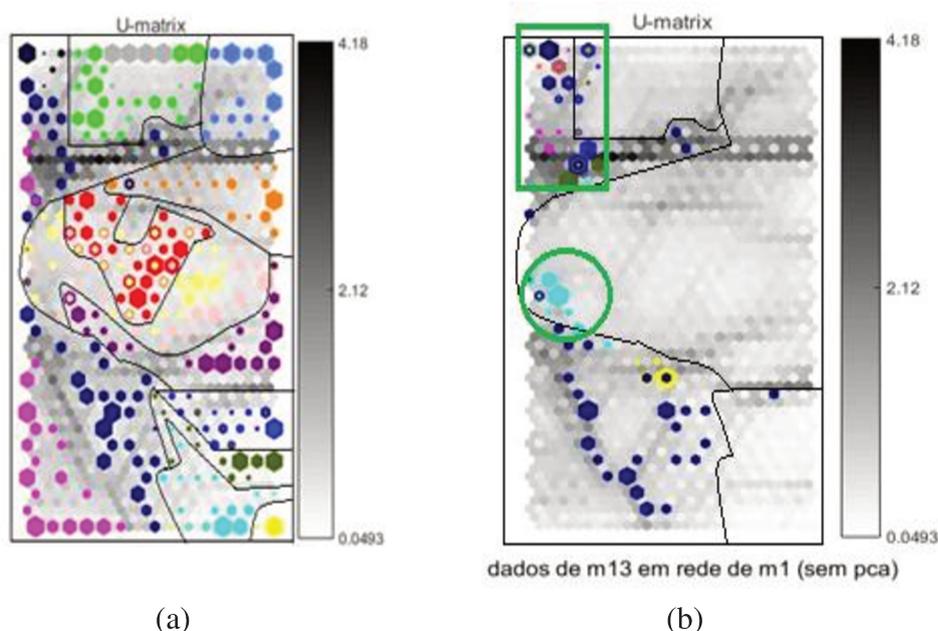


Figura E.2 Representação da matriz U em rede de m1 sem compressão. (a) neurônios ativados pelos dados de treinamento (figura repetida); (b) neurônios ativados pelos dados do m13, extrapolação

Novamente, representação das cores é a mesma dos gráficos anteriores das matrizes U. A figura da esquerda é repetida da Figura 4.29 (a) para efeitos de comparação com a figura (b). Verifica-se que as maiores concordâncias são dos experimentos sem vazamento com batidas na tubulação, em azul escuro.

Os dados de vazamentos capturados pelo microfone distante, m13, ativaram, em sua maioria, no retângulo verde na figura (b), dados sem vazamentos. Neste mesmo retângulo, evidencia-se que hexágonos azuis (dados de vazamento de 2 mm) ativam também neurônios de vazamentos pequenos de 0,5 mm (em comparação com a figura (a) discutida na seção 4 deste trabalho).

Além, evidenciado pelo círculo verde, hexágonos em ciano (ativados por dados de vazamento de 3 mm). Está área, de acordo com treinamento, é de reconhecimento de dados de vazamento de vazamentos de 1 mm simultaneamente com de 0,5 mm.

Nesta extrapolação, em resumo, muitas ocorrências de vazamento serão confundidas com ausência de vazamento, com ou sem batidas. Em poucas ocasiões de vazamento, será reconhecido que ocorre perda de material, sem assertividades quanto à dimensão do vazamento ou à quantidade de vazamentos em ocorrência.

Dados de Teste de m1 em Rede de Treino com Dados de m13

Nesta extrapolação, a rede foi treinada com dados do microfone distante em relação aos vazamentos simulados, m13, e os dados de teste são do microfone próximo aos vazamentos m1.

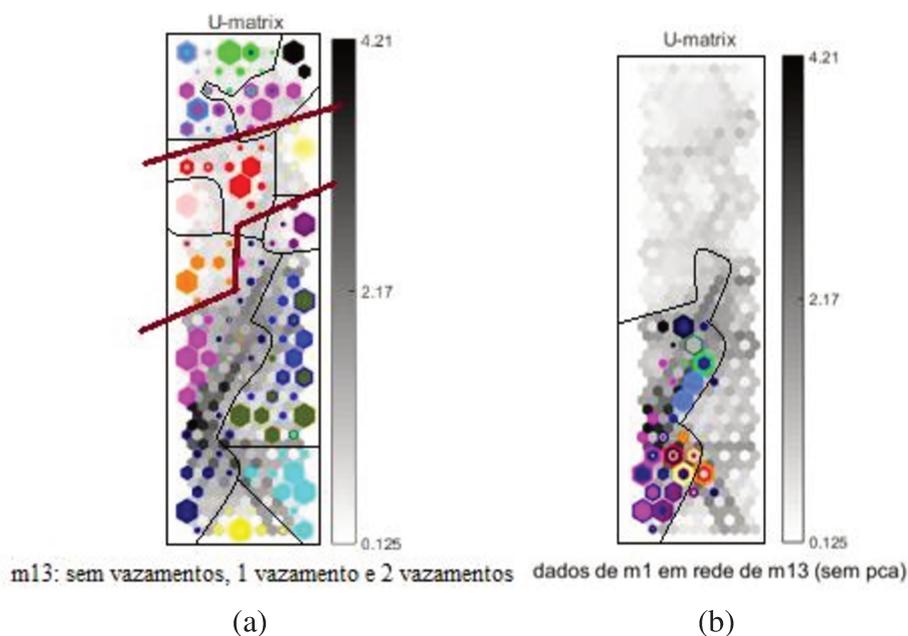


Figura E.3 Representação da matriz U em rede de m13 sem compressão. (a) neurônios ativados pelos dados de treinamento (figura repetida); (b) neurônios ativados pelos dados do m1, extrapolação

Na Figura E.3, à esquerda, a matriz U é também repetida (da Figura 4.30 (a)) e está representada para fins de facilitar a comparação com o teste de extrapolação à direita. A área destacada na figura (b) segue a área equivalente na figura (a).

Percebe-se que praticamente todos os neurônios ativados pelo teste de dados de extrapolação são relativos a situações de sem vazamento com batidas (correspondentes a hexágonos em rosa escuro e azul escuro na figura (a)). Uma explicação para isso pode ser o fato que as situações de batidas são as com maiores perturbações para a resposta sonora ao microfone 13, conforme análise do Apêndice C. Assim, as grandes variações de amplitudes capturadas por m1 de situações experimentais encontram maiores similaridades em situações de batidas por m13.

Como conclusão, este apêndice mostra situações em que, quando a resposta da rede acusa eventos sem vazamentos enquanto há, ou o contrário, as extrapolações

seriam percebidas pela comparação com outros dados acústicos, como o comportamento das amplitudes no tempo e a resposta poderia ser descartada conforme necessário.