



# UFBA

UNIVERSIDADE FEDERAL DA BAHIA  
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS GRADUAÇÃO EM  
ENGENHARIA INDUSTRIAL - PEI

MESTRADO EM ENGENHARIA INDUSTRIAL

MARIZA SOUZA DE MELLO

RECONHECIMENTO DE PADRÕES DE USO DE  
EQUIPAMENTOS HIDRÁULICOS DOMICILIARES

**PEI**

Programa de Pós-Graduação em Engenharia Industrial

**SALVADOR**  
**2018**



**UNIVERSIDADE FEDERAL DA BAHIA**  
**ESCOLA POLITÉCNICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA INDUSTRIAL**

**MARIZA SOUZA DE MELLO**

**RECONHECIMENTO DE PADRÕES DE USO DE EQUIPAMENTOS  
HIDRÁULICOS DOMICILIARES**

Salvador

2018

**MARIZA SOUZA DE MELLO**

**RECONHECIMENTO DE PADRÕES DE USO DE EQUIPAMENTOS  
HIDRÁULICOS DOMICILIARES**

Dissertação de Mestrado Profissional apresentada ao Programa de Pós-graduação em Engenharia Industrial, da Universidade Federal da Bahia, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Industrial.

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Karla Patricia Santos Oliveira Rodríguez Esquerre

Salvador

2018

Mello, Mariza Souza de

Reconhecimento de padrões de uso de equipamentos hidráulicos domiciliares / Mariza Souza de Mello. -- Salvador, 2018.

85 f. : il

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Karla Patricia Santos Oliveira Rodríguez Esquerre.

Dissertação (Mestrado - Programa de Pós-graduação em Engenharia Industrial) -- Universidade Federal da Bahia, Escola Politécnica, 2018.

1. Água. 2. Consumo residencial. 3. Medidor inteligente. 4. Reconhecimento de padrão. I. Esquerre, Karla Patricia Santos Oliveira Rodríguez. II. Título.

**RECONHECIMENTO DE PADRÕES DE USO DE EQUIPAMENTOS  
HIDRÁULICOS DOMICILIARES**

**MARIZA SOUZA DE MELLO**

Dissertação submetida ao corpo docente do Programa de Pós-graduação em Engenharia Industrial da Universidade Federal da Bahia como parte dos requisitos necessários para a obtenção do grau de mestre em Engenharia Industrial.

Examinada por:

Prof. Dra. Karla Patrícia S. O. Rodriguez Esquerre Karla Patricia Rodriguez Esquerre  
Doutora em Engenharia Química, pela Universidade de Campinas, Brasil, 2003

Prof. Dr. Asher Kiperstok Asher Kiperstok  
PhD em Engenharia Química/Tecnologias Ambientais, pela Universidade de Manchester, Inglaterra, 1996

Prof. Dr. Adonias Magdiel Silva Ferreira Adonias Magdiel Silva Ferreira  
Doutor em Engenharia Industrial, pela Universidade Federal da Bahia, Brasil 2015

Salvador, BA - BRASIL  
Julho/2018

## **AGRADECIMENTOS**

Neste trecho, gostaria de expressar minha gratidão diante do apoio, orientação e colaboração de pessoas, sem as quais não me seria possível chegar até aqui. Tenho a certeza, porém, que a recordação e a influência de todos vocês sempre farão parte das vitórias que lograrei alcançar.

Enfatizo minha gratidão a Deus que me aponta caminhos e à minha família que me incentiva a segui-los;

À Profa. Dra. Karla Esquerre, obrigada pela confiança em mim depositada, propiciando-me a oportunidade de trabalhar neste projeto, instrumento de ampliação de meu escopo de conhecimento;

Ao Prof. Dr. Asher Kiperstok pela confiança ao ceder os dados, que foram fundamentais nessa pesquisa;

A CNPQ e FINEP pelo financiamento que possibilitou a realização das pesquisas CT-HIDRO e PROSAB;

A Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) pela bolsa concedida;

A Adelmo Filho, obrigada pelo paciente suporte às minhas iniciais dificuldades e providencial ajuda na compreensão do problema e no tratamento dos dados;

A Profa. Ma. Gabriella Botelho, obrigada pelo suporte ao explicar sobre os detalhes dos projetos, os quais me ajudaram sobremaneira na interpretação dos resultados;

Aos Componentes do grupo GAMMA, obrigada pelo carinhoso acolhimento e indispensável apoio;

Ao Corpo técnico do PEI, Tatiane Woytysiak, Robinson Xavier, Tamiles Bispo e os estagiários, obrigada pela fundamental colaboração ao longo do processo rumo ao meu objetivo;

Aos Amigos, cuja convivência proporcionou-me boa companhia, oportuna ajuda e bons momentos, obrigada.

MELLO, Mariza Souza de. **Reconhecimento de padrões de uso de equipamentos hidráulicos domiciliares**. 2018. 85 f. il. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-graduação em Engenharia Industrial, Escola Politécnica, Universidade Federal da Bahia, Salvador, 2018.

## RESUMO

Modelar o consumo residencial de água é um passo importante para o conhecimento sobre o comportamento dos usuários e, posteriormente, para a proposição de metas realistas que levem ao consumo consciente. As informações disponibilizadas pelos medidores inteligentes facultaram a possibilidade de detecção do equipamento acionado, através de medidas das vazões em intervalos de tempos da ordem de segundos. Dessa forma, o objetivo deste projeto é desenvolver uma ferramenta de identificação de uso de equipamentos, composta do algoritmo de classificação e do algoritmo de identificação e separação de usos simultâneos. No quesito classificação, o algoritmo baseou-se no conceito do *k-nearest neighbor*, avaliando a medida de (dis)similaridade que melhor se adequava aos dados, dentre as DTW, EDR, ERP e LCSS. Os indicadores de qualidade adotados foram a sensibilidade – que indica o percentual dos dados originais classificados corretamente – e a precisão – que avalia o classificador ao apresentar o percentual dos dados assinalados corretamente. Ambos foram calculados tanto em relação a quantidade de eventos quanto ao volume consumido. A medida ERP foi eleita em detrimento das outras pelos valores alcançados pelos indicadores e pela menor produção de resultados com empates. Em termos de sensibilidade com relação ao volume, a bacia sanitária, chuveiro e torneira interna obtiveram respectivamente médias de 71,1%, 67,3% e 78,2%, sendo que para mais da metade das residências conseguiu-se alcançar mais de 70% de detecção entre esses equipamentos. Para torneira externa esse indicador obteve valores abaixo do esperado com média de 26,6%, por conta principalmente de ser confundido com a torneira interna. Em termos de precisão com relação ao número de eventos, o classificador conseguiu para a bacia sanitária, chuveiro, torneira externa e torneira interna médias de respectivamente 72,2%, 51,1%, 59,0%, 88,4%; em algumas casas a capacidade de detecção para a torneira externa chegou a mais de 90%. O algoritmo de identificação e separação de usos simultâneos foi capaz de diferenciar usos particulares de usos simultâneos, além realizar separações, obtendo uma média 83% de acerto entre todas as residências. A integração desses algoritmos foi testada em uma das casas e obteve-se ao final 78% dos perfis em correspondência correta. Esses resultados indicaram que a ferramenta proposta mostrou potencial de ser utilizada, tanto por conta dos resultados obtidos quanto pelo fato de não haver intervenção do pesquisador ao longo do processo.

**Palavras-chave:** água, consumo residencial, medidor inteligente, reconhecimento de padrão.

MELLO, Mariza Souza de. **Reconhecimento de padrões de uso de equipamentos hidráulicos domiciliares**. 2018. 85 f. il. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-graduação em Engenharia Industrial, Escola Politécnica, Universidade Federal da Bahia, Salvador, 2018.

## ABSTRACT

Modeling residential water consumption is an important step towards knowledge about users behavior and, later, for the proposition of realistic goals that lead to the conscious consumption. The information provided by the smart meters enabled the detection of the equipment triggered, through flow measurements in short time intervals. Thus, the objective of this project is to develop an algorithm to identify the use of devices, with both pattern classification, as well as identification and separation of simultaneous uses. The classifier was built based on k-nearest neighbor concept, including the evaluation of the similarity measure that best suited the data among DTW, EDR, ERP and LCSS. The measures adopted were sensitivity - which indicates the percentage of the total events correctly classified - and precision - which indicates the percentage of the data correctly assigned by the classifier. They were calculated based on number of events and volume consumed. The ERP measure was chosen as the one that reached the best performance measures values and that produced the smaller percentage of tie results. The sensitivity average, in terms of volume, for the toilet, shower and internal tap were 71.1%, 67.3% and 78.2% respectively, and for more than half of the residences, it was possible to achieve more than 70% detection between these devices. For external tap the performance measure value was below-expected with an average of 26.6%, mainly due to confusion with the internal tap. In terms of accuracy with regard to the number of events, the classifier achieved for the toilet, shower, external tap and internal tap average of respectively 72.2%, 51.1%, 59.0% and 88.4%; in some homes, the detection capacity for the external tap reached more than 90%. The algorithm of identification and separation of simultaneous uses was able to differentiate particular uses from simultaneous uses and realize separations, obtaining an average 83% of correct outcomes among all the residences. The integration of these sections was tested in one of the houses and 78% of the water flows were correctly matched. These results indicated that the proposed tool showed potential to be used, both because of the interesting results obtained and the fact that there was no intervention of the researcher throughout the process.

**Keywords:** water, residential consumption, smart meter, pattern recognition

## LISTA DE ILUSTRAÇÕES

<i>Figura 1</i> - Esquema explicativo das variações de modelo propostas. O modelo sem restrição (MSR) e modelo com restrição (MCR) em que os equipamentos mais prováveis são pré-selecionados de acordo com a duração e volume do perfil teste. ....	28
<i>Figura 2</i> - Exemplo de resposta de um medidor inteligente acoplado com um data logger, com vazão em l/10s a partir da conversão dos pulsos. ....	31
<i>Figura 3</i> - Exemplos de perfis de uso da torneira interna. ....	36
<i>Figura 4</i> - Exemplos de perfis de uso da torneira externa. ....	37
<i>Figura 5</i> - Exemplos de perfis de uso de chuveiro. ....	37
<i>Figura 6</i> - Exemplos de perfis de uso da bacia sanitária. ....	38
<i>Figura 7</i> – Percentual de uso dos equipamentos em termos de quantidade de eventos (a) e de volume consumido (b), por casa. ....	39
<i>Figura 8</i> - Histogramas das vazões por aparelho, para as residências A, B, C e D. ...	40
<i>Figura 9</i> - Histogramas das vazões por aparelho, para as residências E, F, G e H. ...	41
<i>Figura 10</i> - Histogramas das vazões por aparelho, para as residências I e J. ....	41
<i>Figura 11</i> – Boxplot da duração dos equipamentos por casa. ....	42
<i>Figura 12</i> - Boxplot do volume dos equipamentos por casa. ....	42
<i>Figura 13</i> - Fluxograma da etapa de análise de simultaneidade. S significa simultâneo e NS não simultâneo. ....	46
<i>Figura 14</i> - Fluxograma simplificado da etapa de corte vertical. A letra j representa a posição do vetor X de comprimento N. ....	47
<i>Figura 15</i> - Exemplos de cortes verticais realizados. Há um perfil por gráfico (estado bruto) e a divisão por cores em cada um identifica como o algoritmo atuou. ....	47
<i>Figura 16</i> - Fluxograma simplificado da etapa de corte horizontal. ....	48
<i>Figura 17</i> - Exemplos de corte horizontal. Há um perfil (estado bruto) por gráfico e a divisão por cores em cada um identifica como o algoritmo atuou. O perfil base é o vermelho e em verde e azul estão as seções que foram cortadas. ....	49
<i>Figura 18</i> - Esquema ilustrativo do alinhamento entre séries temporais, a partir do uso do DTW. Na letra A são exibidas as séries Q e C; Na letra B é apresentada a matriz de distorção, com cada elemento sendo $d(x_i, y_j)$ e ainda o caminho (W) que oferece distância mínima entre as séries; e na letra C as séries são alinhadas de acordo com o caminho W (imagem de KEOGH et al., 2005). ....	52

<b>Figura 19</b> – Esquema ilustrativo da restrição da janela por Sakoe e Chiba (1978)...	56
<b>Figura 20</b> - Imagem ilustrativa do emprego do k-NN. O triângulo laranja é o objeto a ser classificado. Para $k=1$ , ele corresponde a classe A. Para $k=3$ , ele também corresponde a classe A, por essa ser a classe dominante (2 de A versus 1 de B).....	57
<b>Figura 21</b> – Fluxograma simplificado dos modelos MCR (a) e MSR (b).....	58
<b>Figura 22</b> –Matriz de confusão ilustrativa, que situa as informações utilizadas para cálculo dos indicadores. A exemplo, nessa estão os nove resultados possíveis em uma tabela de contingência $3 \times 3$ . .....	59
<b>Figura 23</b> - Fluxograma da hipótese 1 (a) iniciando com a classificação e da hipótese 2 (b) iniciando com a análise e separação de perfis simultâneos.....	60
<b>Figura 24</b> - Gráfico de dispersão da diferença entre o volume do perfil pelo algoritmo e pelo grupo (eixo x) versus a diferença entre a duração do perfil pelo algoritmo e pelo grupo, por residência. ....	63
<b>Figura 25</b> – Resultados globais de classificação, apresentando a acurácia para resultados com e sem empate, assim como o percentual de perfis com empate de resultados. ....	65
<b>Figura 26</b> - Mediana da sensibilidade para o volume (a) e mediana da precisão para volume V e número de eventos Q (b).....	72
<b>Figura 27</b> - Esquema de instalação do novo sistema de medição em uma das casas em estudo (BOTELHO, 2013). .....	83
<b>Figura 28</b> – Gráfico boxplot da moda (L/min) para os equipamentos por residência. ....	83
<b>Figura 29</b> - Gráfico boxplot da vazão média (L/min) para os equipamentos por residência. ....	84
<b>Figura 30</b> - Gráfico boxplot da vazão máxima (L/min) para os equipamentos por residência. ....	84

## LISTA DE TABELAS

<i>Tabela 1 - Informações gerais sobre cada residência. O projeto de sigla P1 é o do CT-Hidro e P2 PROSAB.</i> .....	32
<i>Tabela 2 - Distribuição percentual dos acionamentos por equipamento por residência no período investigativo.</i> .....	36
<i>Tabela 3 - Percentual médio de acerto das configurações sugeridas.</i> .....	61
<i>Tabela 4 - Resultado médio do percentual de acerto com as variações de parâmetros.</i> .....	62
<i>Tabela 5 - Resultados do processo de separação, com informação do percentual de acerto e de erros, separados em duas categorias: incompatibilidade horária e diferenças entre duração ou volume.</i> .....	63
<i>Tabela 6 - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a bacia sanitária. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.</i> .....	66
<i>Tabela 7 - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a bacia sanitária. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.</i> .....	66
<i>Tabela 8 - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para o chuveiro. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.</i> .....	67
<i>Tabela 9 - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para o chuveiro. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.</i> .....	68
<i>Tabela 10 - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira externa. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.</i> .....	68

<b>Tabela 11</b> - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira externa. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.....	69
<b>Tabela 12</b> - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira interna. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores. ....	69
<b>Tabela 13</b> - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira interna. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.....	70
<b>Tabela 14</b> - Percentual de perfis com empates para análise inicial e para a parte majoritária dos dados.....	72
<b>Tabela 15</b> - Sumário dos resultados de precisão por volume (V) e quantidade de eventos (Q) para todas as residências utilizando o modelo MSR e a medida ERP. ....	73
<b>Tabela 16</b> - Sumário dos resultados de precisão por volume (V) e quantidade de eventos (Q) para todas as residências utilizando o modelo MSR e a medida ERP. ....	73
<b>Tabela 17</b> – Resultados acurácia dos modelos sugeridos. ....	75
<b>Tabela 18</b> - Matriz de confusão para a hipótese 1. ....	75
<b>Tabela 19</b> - Matriz de confusão para a hipótese 2. ....	75

## LISTA DE QUADROS

<i>Quadro 1 - Relação de trabalhos que lidaram com reconhecimento de padrões de uso de equipamentos. ....</i>	<i>25</i>
<i>Quadro 2 - Informações importantes sobre os modelos avaliados como referência. ...</i>	<i>30</i>
<i>Quadro 3 - Paralelo entre o algoritmo proposto e o de Nguyen et al., (2013b). ....</i>	<i>44</i>
<i>Quadro 4 – Configurações testadas para determinar algoritmo final de identificação e separação de usos paralelos. ....</i>	<i>50</i>
<i>Quadro 5 - Resumo de informações sobre as medidas de (dis)similaridade. ....</i>	<i>53</i>
<i>Quadro 6 - Relação dos templates aplicados por residência. Ao lado de cada template consta o percentual de vezes que foi aplicado. ....</i>	<i>85</i>

## LISTA DE ABREVIATURAS E SIGLAS

1NN	<i>1-Nearest Neighbor</i>
ASP0	Restrição local assimétrica de Sake-Chiba para $P=0$
DTW	<i>Dinamic time warping</i>
EDR	<i>Edit distance on real sequences</i>
ERP	<i>Edit distance with real penalty</i>
KNN	<i>k-Nearest Neighbor</i>
LCSS	<i>Longest common subsequence</i>
MCR	Modelo com restrição
MSR	Modelo sem restrição
pma	Percentual mínimo de aceitação
SP0	Restrição local simétrica de Sake-Chiba para $P=0$

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	16
<b>2. OBJETIVO</b> .....	21
2.1. OBJETIVO GERAL.....	21
2.2. OBJETIVOS ESPECÍFICOS .....	21
<b>3. REVISÃO BIBLIOGRÁFICA</b> .....	22
<b>4. ANÁLISE DOS DADOS</b> .....	31
4.1. COLETA DE DADOS .....	31
4.2. CLASSIFICAÇÃO VIA TRACE WIZARD.....	33
4.3. CARACTERIZAÇÃO DOS DADOS .....	35
4.3.1. PADRÕES – DADOS REFERENCIAIS .....	35
4.3.2. DADOS CLASSIFICADOS VIA TRACE WIZARD .....	38
<b>5. MÉTODO</b> .....	43
5.1. IDENTIFICAÇÃO E SEPARAÇÃO DE USOS SIMULTÂNEOS .....	43
5.1.1. FUNDAMENTAÇÃO.....	43
5.1.2. CONSTRUÇÃO DO ALGORITMO .....	45
A. ANÁLISE DA SIMULTANEIDADE.....	45
B. CORTE VERTICAL .....	46
C. CORTE HORIZONTAL .....	48
5.1.3. PARÂMETROS .....	49
5.1.4. ESTUDO DA CONFIGURAÇÃO DO ALGORITMO .....	49
5.2. CLASSIFICAÇÃO .....	50
5.2.1. FUNDAMENTAÇÃO TEÓRICA .....	50
5.2.1.1. MEDIDAS DE (DIS)SIMILARIDADE .....	50
A. DTW .....	50
B. EDR, ERP E LCSS .....	52

C. OTIMIZAÇÃO DE PARÂMETROS .....	54
5.2.1.2. ADAPTAÇÃO DO ALGORITMO K-NEAREST NEIGHBORS .....	56
5.2.2. MODELOS PROPOSTOS .....	57
5.3. ALGORITMO INTEGRADO .....	60
<b>6. RESULTADOS .....</b>	<b>61</b>
6.1. IDENTIFICAÇÃO E SEPARAÇÃO DE USOS SIMULTÂNEOS .....	61
6.2. CLASSIFICAÇÃO .....	63
6.2.1. ANÁLISE POR EQUIPAMENTO .....	65
A. BACIA SANITÁRIA .....	65
B. CHUVEIRO .....	67
C. TORNEIRA EXTERNA.....	68
D. TORNEIRA INTERNA.....	69
6.2.2. ELEIÇÃO DA MEDIDA DE (DIS)SIMILARIDADE IDEAL .....	70
6.3. TÉCNICAS AUXILIARES PARA AMPLIAÇÃO DO DESEMPENHO DO CLASSIFICADOR.....	73
6.4. ALGORITMO INTEGRADO .....	74
<b>7. CONCLUSÃO.....</b>	<b>76</b>
<b>8. SUGESTÃO DE TRABALHOS FUTUROS.....</b>	<b>78</b>
<b>9. PUBLICAÇÕES .....</b>	<b>79</b>
<b>REFERÊNCIAS .....</b>	<b>80</b>
<b>APÊNDICE A .....</b>	<b>83</b>
<b>APÊNDICE B.....</b>	<b>85</b>

## 1. INTRODUÇÃO

O entendimento do modo de consumo de água residencial é fundamental para o desenvolvimento e implementação de estratégias mais assertivas de gestão de demanda. Considerando os atuais padrões e hábitos de consumo, que culminam no desperdício, espera-se um aumento da demanda de água residencial devido ao crescimento urbano mundial (COMINOLA et al., 2015). Além do aumento da população, a poluição da água, o desenvolvimento urbano, a irrigação agrícola, as mudanças climáticas e secas também contribuem para disparidades entre a disponibilidade de fontes de água de qualidade e demanda de consumo (JORGENSEN et al., 2009). Consequentemente, a sociedade está enfrentando problemas de escassez de água em muitas partes do mundo e a informação de como e quando a água é usada pode ajudar a delinear políticas que visam à redução desse consumo (VAŠAK et al., 2015). No caso do Brasil, embora conhecido pela sua grande riqueza hídrica, a distribuição de água doce é heterogênea e o regime fluvial sofre variações ao longo do ano, ligadas ao regime de precipitações (ANA, 2016). Assim, os centros urbanos brasileiros estão susceptíveis a desafios similares aos relacionados na literatura sobre a disponibilidade hídrica. Dessa forma, mostra-se necessário o aperfeiçoamento do estudo sobre o comportamento do consumo d'água, ampliando para uma abordagem que se baseie em informações mais detalhadas, favorecendo o desenvolvimento de políticas públicas de maior efetividade.

Os estudos relacionados à caracterização do uso de água intradomiciliar foram favorecidos pelos dados provenientes de medidores inteligentes (BENNETT et al., 2013; NGUYEN et al., 2014; LIU et al., 2016). Diferindo dos hidrômetros usuais, os medidores inteligentes são dispositivos eletrônicos de alta resolução, capazes de medir o fluxo de água, energia ou gás natural em intervalos de tempos curtos, com emissor de sinal integrado, unidos a equipamentos de captação automática do sinal emitido. Os dados captados, pelo seu detalhamento, tornam-se importante fonte de informações acerca do consumo interno. Nguyen et al. (2013a) e Bennett et al. (2013), com preocupações similares, tiveram seus estudos focados no consumo residencial de água na Austrália (que enfrentava um período de secas) e ressaltaram que o acesso a dados fornecidos por medidores inteligentes foi um facilitador para identificação dos equipamentos hidráulicos acionados.

Os dados provenientes de medidores inteligentes com considerável nível de detalhamento (por exemplo, com intervalos de medições em ordem de segundos) subsidiam várias vertentes de estudos como, particularmente, a identificação de aparelhos acionados. A ferramenta necessária para identificação de equipamentos demanda uma estrutura que contempla etapas fundamentais como classificação e separação de usos simultâneos. O intuito do primeiro é identificar qual aparelho foi acionado com base nas séries temporais detectadas e, para tanto, demanda o uso de técnicas de reconhecimento de padrões. O segundo tem a função de analisar e dissociar usos simultâneos de água, ou seja, séries temporais que apresentem usos de equipamentos em paralelo. A detecção de vazamentos também é uma etapa importante nesse processo. Ela é bastante útil para o usuário ao indicar possíveis perdas e requer que a sinalização seja realizada em tempo real. Todavia, essa etapa não foi desenvolvida nesse trabalho.

Em termos de gestão dos recursos hídricos por parte das concessionárias, esse tipo de informação aprimora a atuação da gestão da demanda – mudança de paradigma que apoia o uso consciente em detrimento do fornecimento sem critério (STUDART e CAMPOS, 2001). Segundo Liu et al. (2016), embora campanhas de estímulo a conservação da água já sejam utilizadas, o *feedback* à sociedade sobre o consumo detalhado d'água é mais efetivo e pode promover conservação a nível individual e residencial. Eles propuseram um aplicativo de interface amigável, chamado HWU (*Home Water Update*) no qual é possível visualizar informações como porcentagem de consumo do tipo de aparelho pelo total volume d'água, tempo e volume médios de uso por tipo de equipamento, comparar o consumo médio da residência com outra(s) dentro do mesmo bairro (com características parecidas como o número de moradores) e, baseado no setor que mais consome água, ainda fornecer sugestões personalizadas de como reduzir o consumo. Vašak et al. (2015), com o COP (Customer On-line Portal), apresentaram os resultados num formato mais simples com informações sobre frequência de uso por aparelho e o volume mensal médio por aparelho. Nguyen et al. (2014) apresentaram uma proposta similar a Liu et al. (2016), com a adição de informação a respeito do valor associado ao consumo e ainda uma versão para as concessionárias que deveria incluir médias de consumo por localidades, padrões demanda diurna diária, relatórios de previsão de demanda de água para regiões selecionadas com base em dados fornecidos em tempo real. Eles ainda ressaltaram que tais informações poderiam auxiliar também na modelagem de redes de tubulações otimizadas.

Para esse projeto, os dados foram concedidos pelos projetos "Gestão da Demanda por Água em Residências de Baixa e Média Renda – BA" (GESÁGUA) atendendo a Chamada Pública Edital MCT/CNPq/CT-Hidro nº 21/2009, coordenado pela Prof.<sup>a</sup> Dra. Karla Oliveira-Esquerre, e “Uso Racional de Água e Eficiência Energética em Habitações de Interesse Social” atendendo a Chamada Pública: SANEAMENTO AMBIENTAL E HABITAÇÃO 07/2009 – FINEP/PROSAB - Transversal - REDE22, coordenado pelo Prof.<sup>o</sup> Dr. Asher Kiperstok. Ambos os projetos foram realizados na cidade de Salvador, nos bairros Chapada do Rio Vermelho e Plataforma respectivamente e com dados obtidos através da instalação de medidor inteligente e *data logger* por casa. Eles optaram por estudar residências de baixa renda pelo fato das mesmas representarem uma parcela expressiva da população de Salvador, no aspecto socioeconômico. O objetivo de ambos foi estudar o comportamento de consumo das residências utilizando o *software* Trace Wizard®, ferramenta que tem a função de processar os dados advindos dos medidores, extraíndo as séries temporais e associando-as a equipamentos existentes nas residências.

Contudo, embora essa ferramenta tenha as funções de classificação e separação de usos simultâneos, a mesma também é conhecida por requerer intervenção por parte dos pesquisadores, tornando o processo lento e susceptível a erros humanos (MORRISON e FRIEDLER, 2014; COMINOLA et al., 2015). Desta maneira, nesta pesquisa propôs-se uma nova ferramenta de identificação de aparelhos acionados, que atuasse com mínima intervenção dos pesquisadores. Além da questão da intervenção, no corpo de residências estudadas houve predomínio de aparelhos de uso manual em detrimento dos automáticos. Essa particularidade tornou este estudo mais complexo, porque as características de consumo por equipamento são ditadas pelos moradores, que por sua vez são influenciados por questões como cultura local, modo de vida e condições dos equipamentos e instalações residenciais. Isso também pode levar à maior probabilidade de classificações errôneas, a partir do momento em que equipamentos diferentes podem passar a ter comportamentos em comum. Morrison et al. (2014) afirmam que o *software* Trace Wizard® é mais adequado para identificar eventos automatizados como bacias sanitárias, máquinas de lavar louça, máquinas de lavar roupa por conta do comportamento consistente das vazões. Assim, também se buscou criar um modelo de classificação que pudesse abarcar a variabilidade inerente.

Dentre as técnicas de classificação utilizadas nesse campo de estudo, tanto por *softwares* comerciais quanto por outros grupos de pesquisa, há duas vertentes: aquelas

que lidam com séries temporais e as que usam informações derivadas dessas séries como, por exemplo, duração do consumo e volume gasto, vazão máxima verificada, vazão média, moda e seu percentual de ocorrência, entre outros. No primeiro caso, reparou-se o uso do modelo oculto de Markov (NGUYEN et al., 2013a). Todavia pelas suas características, se faz necessário que as séries por equipamento apresentem tendência homogênea e, nesse caso, esse atributo não é predominante. Por outro lado, modelos que lidam com características extraídas das séries temporais como árvore de eventos, redes neurais artificiais, *support vector machine*, *k-nearest neighbors*, modelos lineares robustos, distribuição gaussiana multivariada (Trace Wizard®, Identiflow®, WONDERS et al., 2016; VASAK et al., 2015; FONTDECABA et al., 2013) requerem que esses atributos sejam singulares por equipamento para que seja possível a separação no espaço. No entanto, como citado anteriormente, não é possível garantir uniformidade por equipamento. Avaliando trabalhos sobre classificação de séries temporais em outros temas, notou-se modelos baseados em medidas de (dis)similaridade. Estes são fundamentados na comparação entre uma série temporal desconhecida e séries padrão (conjuntos de séries temporais que representam os comportamentos de cada equipamento) para determinação da classe correta. Esse formato é bastante interessante por ser capaz assimilar variados usos relativos a um mesmo equipamento.

Em meio às medidas de (dis)similaridade conhecidas, destacam-se os *dynamic time warping* (DTW), *longest common subsequence* (LCSS), *edit distance with real penalty* (ERP) e *edit distance on real sequences* (EDR), então foi proposto o estudo avaliativo do potencial dessas medidas para escolha daquela capaz de fornecer o melhor desempenho de classificação. A estrutura do algoritmo proposto foi baseada no conceito do algoritmo *k-nearest neighbors*, cuja ideia principal é determinar a classe do objeto de acordo com seus *k* vizinhos mais próximos. Nesse caso, o objeto é a série temporal desconhecida, cuja distância entre as séries padrão é calculada por uma medida de (dis)similaridade, e ele será classificado de acordo com a classe majoritária entre os *k* vizinhos considerados (WONDERS et al., 2016).

O algoritmo de separação de usos simultâneos, embora bastante importante, de modo geral não é explorado pelos pesquisadores. Há um consenso entre eles de que essa etapa envolve desafios e limitações que dificultam a criação de um modelo capaz de fazer divisões assertivas, logo poucos trabalhos a esse respeito estão disponíveis. Os *softwares* Trace Wizard® e Identiflow® têm a opção de averiguação de usos paralelos de

equipamentos, mas não informam quais as principais técnicas utilizadas para tanto. Por outro lado, Nguyen et al. (2013b) desenvolveram um algoritmo cujo modelo fundamentou-se numa varredura do vetor gradiente em busca de verificar comportamentos anômalos de acordo com premissas estabelecidas. Similarmente, essa técnica foi utilizada nesta pesquisa, em conjunto com estatística descritiva e premissas determinadas de acordo com os comportamentos visualizados no banco de dados.

## 2. OBJETIVO

### 2.1. OBJETIVO GERAL

O principal objetivo deste trabalho foi desenvolver uma ferramenta de reconhecimento do uso de equipamentos hidráulicos, a partir de dados provenientes de medidores inteligentes, composta dos algoritmos de classificação e de separação de usos simultâneos.

### 2.2. OBJETIVOS ESPECÍFICOS

- Agregar ao algoritmo de classificação o aspecto de variabilidade de usos, que é característica real do problema;
- Buscar técnicas auxiliares adicionais para aperfeiçoar o desempenho do classificador;
- Construir um algoritmo de identificação e separação de usos simultâneos que seja adaptável a dados provenientes de medidores com outras resoluções e intervalos de medição;
- Estudar a melhor disposição entre os algoritmos de classificação e separação de usos simultâneos;
- Avaliar a necessidade de intervenção do pesquisador frente à qualidade da resposta da ferramenta;
- Auxiliar na difusão de métodos para trabalhar dados de medidores inteligentes, a partir dos resultados obtidos nesse trabalho e nas publicações advindas dele.

### 3. REVISÃO BIBLIOGRÁFICA

Em termos de reconhecimento de padrões, Cominola et al. (2015) e Morrison e Friedler (2014) realizaram interessante revisão e análise das abordagens mais utilizadas a partir dos dados de medidores inteligentes. Morrison e Friedler (2014) criticaram e compararam seis ferramentas distintas, dentre eles três que lidam com identificação de padrões: Trace Wizard® e Identiflow®, ambos comerciais, baseados em árvore de decisão; e HydroSense®, também comercial, baseado num modelo de classificação probabilístico. Cominola et al. (2015) revisaram 134 trabalhos, categorizando-os entre quatro contribuições distintas, dentre elas, a categorização do uso final da água, avaliando as mesmas ferramentas que Morrison e Friedler (2014). Nguyen et al. (2011, 2013a, 2013b, 2014) apresentaram um *software* em desenvolvimento, Autoflow, baseado principalmente no modelo oculto de Markov para classificação, além de outras técnicas auxiliares para melhoria do desempenho.

O Trace Wizard® foi desenvolvido em 1994 pela Aquacraft Inc. e nele a classificação do evento é determinada via uso de árvore de decisão, a qual avalia a similaridade através de parâmetros pré-definidos manualmente para cada equipamento (a estrutura do processo pode ser lida em Cominola et al. (2015)). Após o processo de classificação, o analista posteriormente deve conferir se a mesma foi realizada corretamente. A ferramenta ainda é capaz de separar eventos simultâneos horizontalmente, ou seja, separar um evento base (maior preponderância) de eventos que ocorram no mesmo período, apresentados como picos anormais de uso; também tem a possibilidade de separar eventos contíguos, mas demanda da interferência do pesquisador para informar onde um evento termina e o outro começa (DEOREO et al., 2011). Tais intervenções adicionam à ferramenta necessidade considerável de interferência humana, o que torna o processo global lento e fortemente dependente da experiência do analista na compreensão do comportamento das séries temporais (COMINOLA et al., 2015). Embora haja a facilidade com relação à compatibilidade com qualquer *data logger*, o *software* é mais adequado para identificar aparelhos automatizados como bacias sanitárias, máquinas de lavar louça, máquinas de lavar roupa, temporizadores de irrigação e sistemas de regeneração de tratamento de água, por conta do perfil<sup>1</sup> pouco variável das vazões (MORRISON et al., 2014).

---

<sup>1</sup> A palavra perfil neste trabalho tem o significado de série temporal.

O Identiflow® foi desenvolvido no Reino Unido pelo WRcGroup e, assim como o Trace Wizard®, aplica árvore de decisão para identificar e classificar eventos usando informações discriminantes sobre cada equipamento utilizado na residência. Adicionalmente leva em consideração o tipo de habitação e arranjos de encanamento. Nele, o analista também interfere para revisar e refinar as análises. Além dessa desvantagem, o Identiflow® não está disponível para compra ou aluguel e só é oferecida como parte dos serviços de consultoria WRc no Reino Unido (MORRISON et al., 2014).

Diferentemente dos outros dois, a tecnologia HydroSense® envolve tratamento de dados provenientes de um sensor de pressão instalado num ponto único da casa e é capaz de desagregar<sup>2</sup> e identificar automaticamente eventos de uso de água, com base num modelo de classificação probabilístico. Segundo Morrison et al. (2014), os autores do *software* afirmaram que a tecnologia é sensível o suficiente para determinar o uso de um equipamento individual assim como o uso de água quente ou fria. Ao contrário das outras técnicas, o HydroSense® é automático, ou seja, não requer intervenção do analista, diminuindo a possibilidade de erro humano (MORRISON et al., 2014). Dentre as limitações, Cominola et al. (2015) afirmaram que a ferramenta apresenta baixo desempenho em residências agregadas. Como os autores não se referem a situações em que a pressão na rede de distribuição de água não é estável, não foi possível avaliar os possíveis efeitos dessas variações no desempenho e precisão do HydroSense®.

O SEQREUS (*South East Queensland Residential End Use Study*) foi um projeto desenvolvido com o objetivo de estudar o consumo de água residencial, usando dados de medidores inteligentes e o *software* Trace Wizard®, no sudoeste do estado de Queensland, Austrália. A partir desses dados, Nguyen et al. (2011, 2013a, 2013b, 2014) buscaram desenvolver gradativamente uma ferramenta de identificação de aparelhos acionados e separação de usos paralelos, além de incluir técnicas auxiliares para melhorar a qualidade de desempenho. O nível de detalhamento fornecido foi bastante elucidativo e os artigos são apresentados a seguir.

O início do método consistiu na organização dos padrões de fluxo dos equipamentos em estudo (NGUYEN et al., 2011). Os dados foram provenientes de medidores inteligentes instalados com resolução de 0,0014L/pulso lidos num intervalo de

---

<sup>2</sup> Neste contexto a palavra desagregar tem o significado de quebrar a série total do dia removendo os zeros e separando as séries temporais detectadas pelo medidor de vazão.

5 segundos e seus usos finais<sup>3</sup> separados manualmente pelos analistas via uso do Trace Wizard® (vide Quadro 2). Nessa etapa, perfis de um mesmo equipamento foram reunidos em grupos menores, através do uso do DTW, e de cada um foi escolhido um perfil representante. A diminuição de perfis de fluxo característicos auxilia na redução do tempo computacional em algoritmos de classificação, entretanto os autores informam que o procedimento não é seguro em equipamentos cujos padrões de consumo apresentam grande variabilidade. No processo é possível que determinados perfis sejam bastante similares a uma referência, no entanto diferentes entre si em termos de magnitude e comprimento; assim, no processo de eliminação, perfis podem ser incorretamente eliminados. No âmbito da classificação de usos, os autores construíram um algoritmo baseado no uso do modelo oculto de Markov (*hidden Markov model*, HMM em inglês) em conjunto com características dos usos por equipamento (volume, duração, moda, pico), a medida de dissimilaridade DTW, probabilidade do horário do dia e informações fornecidas pelos moradores. O HMM foi testado em 252 casas e o algoritmo aprimorado, com base em informações complementares, foi implantado em três residências separadamente (NGUYEN et al., 2013a). Os resultados foram sumarizados no Quadro 1 para ilustração do potencial atingido pelo algoritmo proposto, dividido em duas seções: a primeira com o resultado geral, apresentando a acurácia (total de eventos corretamente identificados pelo classificador) em termos de quantidade de eventos (Q) e volume (V); na segunda, os resultados para uma residência desta vez aplicando informações descritivas, probabilidades horárias e o DTW para aperfeiçoar os resultados. Na sequência, foi desenvolvido o algoritmo de separação de eventos combinados através do uso de *gradient vector filtering* (filtragem do vetor gradiente) (NGUYEN et al., 2013b). Por fim, Nguyen et al. (2014) propuseram uma melhoria na ferramenta ao adicionar um algoritmo de auto aprendizado, capaz aprimorar-se através da reavaliação de perfis que não conseguiram ser classificados e agrupando-os em novas categorias. Eles afirmaram que os testes realizados mostraram alta precisão, entre 80 e 90% para a maioria das categorias, no entanto a ferramenta ainda requer intervenção humana para aumentar os níveis de precisão.

Beal et al. (2011), Nguyen et al. (2013a) e Cominola et al. (2015) listaram vários trabalhos (Quadro 1) cujo objetivo foi identificar a utilização de equipamentos hidráulicos

---

<sup>3</sup> O termo uso final está associado ao nível de detalhamento dos dados. Nesse caso, o uso final é o uso do equipamento.

residenciais, indicando a localidade do estudo, a ferramenta utilizada e o número de residências em estudo.

**Quadro 1** - Relação de trabalhos que lidaram com reconhecimento de padrões de uso de equipamentos.

Localização	Ferramenta	Nº de residências	Referência
Austrália	Trace Wizard®	120	Bouwer (2000)
		1500	Beal et al. (2011)
		720	Loh et al. (2003)
		100	Roberts (2005)
		10	Mead (2008)
		10	Mead and Aravinthan (2009)
		200	Willis et al. (2009a)
		151	Willis et al. (2009b)
	252	Willis et al. (2010, 2011a)	
	Autoflow	252	Nguyen et al. (2013a)
3 (de 252)		Nguyen et al. (2013b)	
3 (de 252)		Nguyen et al. (2014)	
Water Use Signature Patterns <sup>a</sup>	11	Cardell-Oliver (2013a)	
	187	Cardell-Oliver (2013b)	
Nova Zelândia	Trace Wizard®	12	Heinrich (2007)
Reino Unido	Identiflow®	250	Kowalski and Marshallsay (2003)
		N/A	Kowalski and Marshallsay (2005)
EUA	Trace Wizard®	16	DeOreo and Mayer (1994)
		16	Mayer and DeOreo (1995)
		N/A	DeOreo et al. (1996)
		10	DeOreo and Mayer (2000)
		37 (de 1188)	Mayer et al. (2000)
		33	Mayer et al. (2003)
		33	Mayer et al. (2004)
		209	Aquacraft Inc. (2011)
1000	DeOreo (2011)		

<sup>a</sup> Essa ferramenta lida com dados de frequência horária.

Além da aplicação de *softwares* comerciais, outros modelos de classificação alternativos foram explorados por outros pesquisadores. Fontdecaba et al. (2013) construíram um algoritmo cujo modelo foi distribuição Gaussiana multivariada, com as variáveis sendo escolhidas através de heurística não aleatória. Os perfis foram divididos em início, meio e fim, e variáveis como duração, volume, vazão e média foram utilizados. Os autores ajustaram o modelo usando conjuntamente os dados de todas as residências e os equipamentos modelados estão listados no Quadro 2. Observaram-se problemas com a variabilidade tanto na unificação dos dados quanto na modelagem por residência, logo os autores optaram pelo uso dos dados unificados. O desempenho do algoritmo de classificação foi avaliado em termos de especificidade (indicador de qualidade com a

mesma definição que a acurácia para Nguyen et al. (2013a)) e da sensibilidade, sendo o percentual total de eventos classificados corretamente. Ambos foram calculados em termos de quantidade de eventos e volume consumido. Vašak et al. (2015) empregaram a técnica de classificação linear robusta do vetor de características (dados no Quadro 2). Eles optaram por utilizar quantização ao nível de 1 pulso por litro em intervalos de 1 segundo e por conta disso houve grandes perdas de informação. Eles fizeram uma pré-remoção de aparelhos menos usados e utilizaram as variáveis duração, volume, vazão máxima, moda e número de ocorrências da moda,  $\sin 2\pi t/T$  e  $\cos 2\pi t/T$  ( $t$  é o número de minutos decorridos a partir da meia-noite e  $T$  o número de minutos em 24 horas), assim como os quadrados dessas variáveis e multiplicação de duplas, totalizando 35 variáveis, para mitigar a não linearidade. Partindo do princípio que equipamentos do mesmo tipo podem apresentar diferentes comportamentos por casa, os autores testaram o algoritmo implementando por residência e reunindo todos os dados. No treinamento, tanto utilizando 7 quanto 35 variáveis, os resultados foram superiores para dados separados por residência; eles também chegaram à conclusão de que o uso de 35 variáveis foi a melhor opção, usando a proporção treinamento/teste de 9/1. Já Wonders et al. (2016) realizaram um estudo comparativo da eficiência de classificação entre três diferentes técnicas de aprendizado de máquinas: Redes neurais artificiais (RNA), *support vector machine* (SVM) e *k-nearest neighbour* (KNN). Eles testaram o aumento do tamanho do banco de dados de treinamento através da produção de dados sintéticos, variaram os parâmetros pertinentes a cada técnica, assim como também variaram a quantidade de variáveis inseridas. O estudo foi realizado com dados do banheiro de uma casa com dois moradores. Os autores justificaram a abordagem do uso de um só cômodo na redução de erros de classificação que são factíveis quando se modela equipamentos distintos, mas que têm comportamentos similares. Os resultados apontaram que o treinamento com o banco de dados estendido foi mais bem-sucedido e a técnica KNN foi a que obteve melhor desempenho.

Comparando os resultados entre esses modelos, nota-se que o do Wonders et al. (2016) obteve o melhor desempenho, no entanto esse resultado foi obtido num sistema menos complexo que os outros, em termos de variedade de equipamentos monitorados. Comparando a acurácia de Vašak et al. (2015) com a especificidade de Fontdecaba et al. (2013), em termos de quantidade de eventos, não é possível observar preponderância entre eles. Notou-se surpreendentemente que a capacidade de detecção da máquina de

lavar pratos do modelo de Fontdecaba et al. (2013) foi baixa, no entanto Nguyen et al. (2013a), que indicaram ter tido problemas semelhantes, apontam que quando muitos programas de um mesmo equipamento são utilizados, a capacidade de detecção fica comprometida.

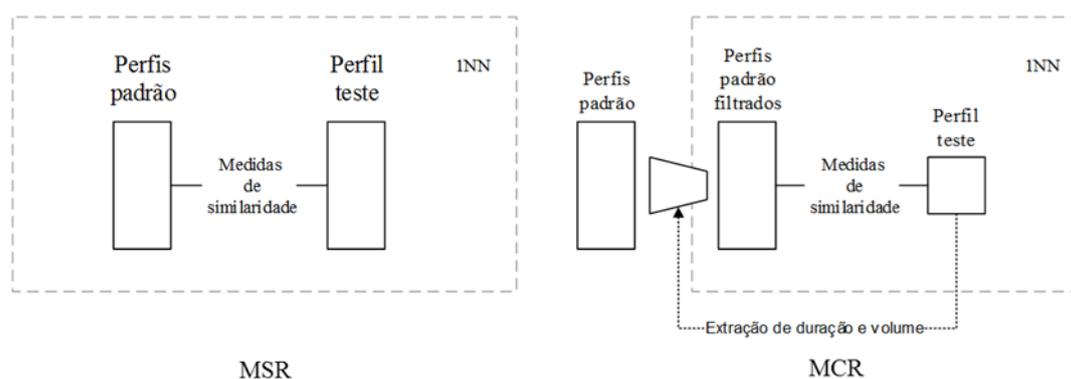
Dentre as técnicas apresentadas, existem limitações para aplicação das mesmas por conta de disponibilidade ou das premissas, como por exemplo:

- O Identiflow® não está disponível para uso fora do Reino Unido;
- O HydroSense® modela dados provenientes de medidores de pressão;
- Os Trace Wizard® não é recomendado para equipamentos que apresentem grande variabilidade de comportamento;
- Não há informação sobre a disponibilidade do Autoflow, porém a principal técnica de classificação é o modelo oculto de Markov que, segundo Nguyen et al. (2013a), é uma técnica poderosa desde que lide com dados que apresentem certa homogeneidade de comportamento, ou seja, mesma limitação que o Trace Wizard®;
- Wonders et al. (2016), Vašak et al. (2015) e Fontdecaba et al. (2013) construíram algoritmos cujos dados de entrada dos modelos são as características dos perfis e por isso também passam pela mesma limitação que os dois anteriores, pois a proximidade de comportamentos entre grupos distintos dificulta a separação dos grupos no espaço.

Neste trabalho, os aparelhos hidráulicos mais utilizados são torneiras (cozinha, banheiros, área externa), chuveiros e bacias sanitárias. Esses tipos de uso dependem do manuseio humano, que está condicionado aos costumes dos usuários, nível de consciência sobre o uso adequado d'água e às condições das instalações hidráulicas. Além disso, por conta da limitação de tipos de aparelhos, é natural assumir que esses acabam acumulando várias funções, contribuindo para aumento da variação de duração e volume dos acionamentos para um mesmo tipo de equipamento.

Dentre as diversas técnicas analisadas referentes à classificação de séries temporais, o uso de medidas de (dis)similaridade se destacou ao ser capaz de considerar a variabilidade existente. A escolha das medidas de (dis)similaridade perpassa pelo conhecimento das características das séries temporais que podem ser geradas. Pela própria ideia de variabilidade, pode-se inferir que as durações de usos variam (perfis com

comprimentos distintos) e, portanto, distâncias como euclidiana ou *Manhattan* não podem ser utilizadas. Por outro lado, as medidas como DTW, LCSS, ERP e EDR cobrem essa limitação, são utilizadas para fins de reconhecimento de padrões e são aplicadas com frequência em trabalhos acadêmicos (LIN et al., 2012; VLACHOS et al., 2002; KEOGH et al., 2005; CHEN et al., 2004; CHEN et al., 2005). A seleção da medida de (dis)similaridade demanda que ela possa captar a dinâmica entre as séries temporais e realizar a classificação de forma coerente. Wang et al (2012) realizaram um trabalho de avaliação de certas técnicas de classificação para séries temporais e, dentre elas, está o uso de medidas de (dis)similaridade aplicadas em 38 bancos de dados reais. Uma das conclusões foi que à medida que o tamanho do banco de dados aumentava, aumentava a convergência da precisão das LCSS, EDR e ERP para a distância euclidiana. Além disso, elas apresentaram precisões muito próximas ao DTW. Esses autores aplicaram essas medidas usando o classificador INN<sup>4</sup> e afirmaram que os maiores benefícios neste tipo de abordagem são a capacidade de avaliar diretamente a eficácia de cada uma das medidas (uma vez que o desempenho de INN depende da medida escolhida) e da técnica não requerer estimativa de parâmetros. Sendo assim, esse algoritmo de classificação foi o adotado neste trabalho, explorando ainda a possibilidade de pré-selecionar os equipamentos para a fase de classificação via uso de restrições de volume e duração, tanto para nortear as comparações quanto para redução do tempo computacional (Figura 1).



**Figura 1** - Esquema explicativo das variações de modelo propostas. O modelo sem restrição (MSR) e modelo com restrição (MCR) em que os equipamentos mais prováveis são pré-selecionados de acordo com a duração e volume do perfil teste.

<sup>4</sup> *k*-nearest neighbors com  $k = 1$

Além do processo de classificação, é imprescindível a análise e separação de perfis que apresentem usos simultâneos de equipamentos. Não raro, numa residência, é perfeitamente possível que dois ou mais aparelhos sejam acionados ao mesmo tempo. Isso causa uma superposição de vazões e necessita de um procedimento de separação para posterior classificação, uma vez que dentro do conjunto de referências somente há registros de equipamentos a nível individual. Essa etapa é complexa, pois não há informações sobre quantos equipamentos estão funcionando em paralelo e nem os tempos de início e término desses eventos (NGUYEN et al., 2013b). Na literatura há poucas referências a cerca de *softwares* ou pesquisadores que tratem desse tipo de análise. Os *softwares* Trace Wizard® e Identiflow® têm a capacidade de efetuar separações, no entanto ambos demandam intervenção do analista e não informam qual a técnica utilizada para separação dos perfis. Já Nguyen et al. (2013b) propuseram o uso de vetor de gradiente filtrado, em conjunto com outras informações não detalhadas pelo autor, para realizar uma varredura no intuito de identificar variações no comportamento da série temporal que correspondam a superposição de usos. O emprego do gradiente do vetor é interessante, e até intuitivo, desde que é necessário avaliar ponto a ponto as variações no mesmo. O algoritmo produzido por eles tem a capacidade de realizar múltiplos cortes dentro do mesmo perfil, tanto horizontal quanto verticalmente; em adição, o mesmo é capaz de realizar varreduras sucessivas, ou seja, realizar um corte e verificar se nele é necessário efetuar outros cortes.

**Quadro 2 - Informações importantes sobre os modelos avaliados como referência.**

Referência	Número de Residências	Período	Local	Medidor	Técnica principal	Técnicas auxiliares	Equipamentos	Indicadores			
								Acurácia		Residência 3	
								Total			
								$V^a$	$Q^b$	HMM (Q)	Modelo aprimorado (Q)
NGUYEN et al. (2013a)	252	2 meses	Austrália	0,014l/s	Modelo oculto de Markov	- DTW - Funções de probabilidade relacionados ao horário do dia - Parâmetros físicos	Bacia sanitária 1	83,80%	71,50%	80,1% <sup>c</sup>	90,3% <sup>c</sup>
							Bacia sanitária 2	82,60%	69,70%	-	-
							Banheira	67,30%	54,20%	42,10%	70,70%
							Chuveiro	74,50%	69,60%	84,30%	94,50%
							Irrigação	73,90%	77,30%	46,20%	66,10%
							Máquina de lavar pratos	84,40%	81,80%	87,10%	96,50%
							Máquina de lavar roupas	85,20%	74,20%	86,60%	93,20%
Torneiras internas <sup>d</sup>	72,90%	73,40%	73,30%	80,10%							
FONTDECABA et al. (2013)	8	3 meses	Espanha	0,1 l/s	Distribuição Gaussiana Multivariada	Heurística não aleatória	Bacia sanitária	76,12%	72,41%	61,44%	62,79%
							Chuveiro	75,36%	49,03%	90,75%	51,94%
							Máquina de lavar pratos	7,73%	9%	56,50%	50%
							Máquina de lavar roupas	74,82%	69,41%	36,86%	30,78%
							Torneiras internas	81,70%	91,37%	42,23%	76,96%
VAŠAK et al. (2015)	74	2 semanas	Inglaterra	1 l/s	Multicategory Robust Linear Programming (MRLP)	Quantização das vazões	Bacia sanitária 1	7 variáveis		35 variáveis	
							Bacia sanitária 2	72,90%		63,50%	
							Chuveiro	69,60%		64,40%	
							Máquina de lavar pratos	79,20%		60,03%	
							Torneiras internas	80,20%		67,50%	
WONDERS et al. (2016)	1	2 meses	Inglaterra	0,001 l/s	Teste de desempenho: RNA, SVM e KNN	Sintetização de dados para adicionar ao treinamento	Bacia sanitária	KNN - Precisão (Q)		KNN - Recall (Q)	
							Bacia sanitária + Chuveiro	64%		100%	
							Bacia sanitária + Lavatório	75%		100%	
							Chuveiro	100%		100%	
							Lavatório	100%		71%	

a: Volume corretamente detectado;

b: Quantidade de eventos corretamente detectada;

c: As bacias foram reunidas;

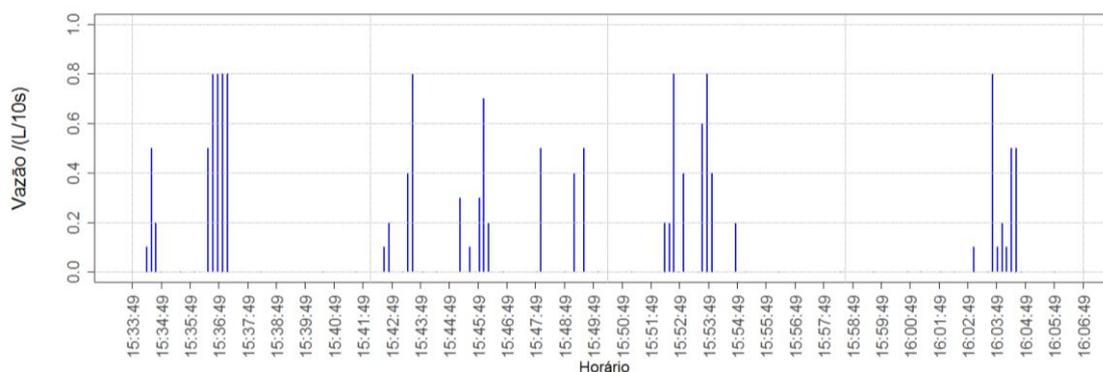
d: Nesses trabalhos listados, exceto WONDERS et al. (2016), as torneiras internas foram reunidas por conta da semelhança de padrão de consumo.

e: Proporção de dados: 90% treinamento e 10% validação.

## 4. ANÁLISE DOS DADOS

### 4.1. COLETA DE DADOS

Os projetos que concederam os bancos de dados utilizados foram apresentados anteriormente. No caso do projeto realizado no bairro Chapada do Rio Vermelho, o período monitoramento foi entre dezembro de 2010 e janeiro de 2012 e analisou o consumo de água intrafamiliar de duas residências. Próximo ao final do período de monitoramento, uma das famílias mudou-se da casa e outra começou a residir. A nova família concordou em continuar a pesquisa e, como seus integrantes apresentaram comportamento distinto de consumo d'água, foram consideradas três residências ao invés de duas. O segundo projeto ocorreu entre dezembro de 2012 e abril de 2014 e foram monitoradas sete residências no bairro de Plataforma, seguindo um método similar ao outro. Em ambos os projetos foram instalados um hidrômetro associado a um *data logger* na saída do reservatório de água (ver Figura 27 no Apêndice A). O hidrômetro instalado (ambos os projetos) foi da marca LAO (Liceu de Artes e Ofício de São Paulo) com vazão nominal de 1,5 m<sup>3</sup>/h, vazão mínima de 0,030 m<sup>3</sup>/h, que emite pulsos a partir de sensores optoeletrônicos com fator K=0,1L/pulso, ou seja, emite um pulso a cada 100mL de fluido que atravessa o hidrômetro. O *data logger*, da marca NOVUS® produtos eletrônicos Ltda, modelo LogBox-DA-64k IP65, foi configurado para armazenar os pulsos emitidos pelo hidrômetro a cada 10s (BOTELHO, 2013) (como exemplificado na Figura 2). Todos os hidrômetros foram calibrados conforme a Portaria do INMETRO n°246:2000 e declarados em conformidade. Em ambos os projetos foi realizada a etapa de avaliação investigativa para estudo dos padrões de consumo por equipamento com ajuda dos residentes.



**Figura 2** - Exemplo de resposta de um medidor inteligente acoplado com um *data logger*, com vazão em L/10s a partir da conversão dos pulsos.

Em geral, os equipamentos comuns entre essas residências foram: bacias sanitárias, chuveiros e torneiras na cozinha, banheiro e área externa. Houve ainda presença de máquinas de lavar e filtro, no entanto, numa frequência menor. No caso das torneiras, houve simplificações de categoria por conta de comportamentos similares. Isso ocorreu em todas as residências para as torneiras da cozinha e do banheiro, categorizadas como torneira interna. As torneiras externas em geral estão associadas a usos mais longos e, portanto, são dissociadas de torneiras internas. Na Tabela 1 constam informações sobre as residências relacionando o projeto, número de moradores, quantidade de pontos de água assim como a quantidade por equipamento, o tempo de monitoramento em dias e o tempo de monitoramento da semana investigativa.

*Tabela 1 - Informações gerais sobre cada residência. O projeto de sigla P1 é o do CT-Hidro e P2 PROSAB.*

Casas		A	B	C	D	E	F	G	H	I	J
Projeto		P1	P2	P1	P2	P1	P2	P2	P2	P2	P2
Número de moradores		2	2	2	2	3	6	2	2	3	2
Número de pontos de água		9	4	7	5	10	9	10	5	6	6
Quantidade de equipamentos por casa	Bacia	2	1	1	1	2	2	2	1	1	2
	Chuveiro	2	1	1	1	2	2	2	1	1	1
	Torneiras internas	3	2	2	2	3	4	3	2	3	2
	Torneira externa	2	0	2	1	2	1	2	1	1	1
	Filtro	0	0	1	0	0	0	0	0	0	0
	Máquina de lavar	0	0	0	0	1	0	1	0	0	0
Tempo de monitoramento (dias)		63	320	305	396	176	236	222	307	341	246
Tempo de monitoramento da semana investigativa (dias)		3	8	11	16	13	8	8	9	14	11

Para fins da construção dos algoritmos, o uso de bacias sanitárias, chuveiros, torneiras internas e torneiras externas foram os usos finais considerados, uma vez que correspondem aos equipamentos hidráulicos com maior frequência em todas as residências simultaneamente. No caso específico das torneiras externas nas casas D, F, G e J, por estarem localizadas em pontos em que a equipe identificou que os equipamentos de medição poderiam ser furtados, optou-se por utilizar formas alternativas para calcular os parâmetros de entrada para o *software*. Assim, não houve séries temporais disponíveis e por essa razão as torneiras externas dessas residências foram excluídas.

#### 4.2. CLASSIFICAÇÃO VIA TRACE WIZARD

O procedimento de aplicação do *software* Trace Wizard® consiste em selecionar os equipamentos a serem modelados, já de posse de suas características para serem inseridas, como duração, volume e moda. Nessa etapa é importante selecionar a ordem dos equipamentos, pois é com base nela que ocorre o processo de classificação. As características do perfil teste são comparadas aos dos equipamentos listados de modo que, se houver compatibilidade, dentro de um limite estabelecido, o perfil é rotulado ao determinado equipamento, sem análise dos posteriores. Logo, é importante ordenar a partir daqueles cujos comportamentos são mais homogêneos até aqueles que não o são (BOTELHO, 2013). Tendo em vista as considerações, as listas com ordem dos equipamentos (*templates*) podem ser observadas no Quadro 6 no Apêndice B. Nele constam os *templates* por casa, sendo que podem haver mais de um por unidade, de acordo com a compreensão do pesquisador e acontecimentos nas residências. Em geral, as ocorrências de vazamento exigiram adaptação do *template* vigente; no projeto CT-Hidro houve etapa de substituição de equipamentos de alto consumo próximo ao final do monitoramento e foram estabelecidos *templates* pós-substituição. Ao lado de cada *template*, há o seu percentual de ocorrência dentro do período de monitoramento; na casa B foram registrados quatro, mas dois deles foram empregados numa pequena porcentagem de dias (4%) e não foram apresentados por não impactarem nas análises. Vale ressaltar que o tratamento dos dados para cada residência foi realizado por membros diferentes da equipe dos projetos. Avaliando a ordenação dos itens dos *templates* foi possível identificar por casa quais aparelhos eram mais comportados e de antemão já apontaram dissemelhanças entre as residências, visto que nenhum *template*, dentre aqueles mais aplicados, foi exatamente igual.

Sobre este último ponto, os relatórios produzidos em ambos os projetos apontaram peculiaridades sobre o consumo d'água, tanto relacionado aos moradores quanto aos equipamentos (BOTELHO, 2011, 2014; BÔAS, 2014; CRUZ, 2014; JESUS, 2014; SANTOS, 2014). Em síntese, o uso de bacia sanitária para a maioria das casas apresentou variações parecidas, no entanto duas delas demonstraram consumo mais elevado por conta de vazamentos, que não foram prontamente consertados (casas B e F). Em residências com o uso de bacias acopladas (casas I e H), o consumo mostrou-se dentro do limite recomendado pela norma NBR 15097-2011.

As torneiras da cozinha e do banheiro, por apresentarem comportamentos parecidos, foram enquadradas na categoria torneiras internas, todavia no projeto CT-Hidro esses equipamentos foram modelados individualmente, não obstante suas características também eram muito parecidas. Em termo de consumo, a casa G apresentou altos valores por conta dos moradores acionarem ao máximo o aparelho; na residência J funcionava concomitantemente um bar, que também eventualmente foi utilizado para a lavagem de pratos e utensílios de cozinha, assim o consumo através desse equipamento foi alto por conta dos processos de cocção e limpeza.

O uso externo de água, através de torneiras, máquina de lavar roupas ou mangueiras, em geral está associado a consumos mais elevados por conta de atividades de lavagem de roupas, da área externa, entretanto nem todas as casas possuíam ponto externo de água. Nestes casos, chuveiros e torneiras internas localizadas na cozinha ou banheiro acumularam funções correlatas às citadas. O consumo em geral também foi afetado pela dinâmica dos moradores, por exemplo, nas residências com 2 moradores em geral um saía para trabalhar e outro fica na casa, daí para o último o consumo estava susceptível às suas escolhas que poderiam sofrer aleatoriedade. Na residência B notou-se aumento no uso de chuveiros e quando a equipe entrevistou o morador, o mesmo informou que a sensação de calor por conta da temperatura ambiente o motivou a refrescar-se mais. Na residência J, a água para a descarga da bacia externa era fornecida via uso de baldes. Na casa A os moradores mostraram certa resistência em cooperação e quando ocorreram problemas com vazamentos, o problema perdurou sem solução.

Essas peculiaridades citadas, além de outras abordadas nos relatórios referentes aos projetos demonstraram que, embora haja proximidade entre as residências (em cada pesquisa) e similaridade no perfil socioeconômico, bem como compartilharem praticamente do mesmo sistema hidráulico-sanitário, cada família tinha a sua particularidade, no que tange o consumo de água, conferidas aos seus hábitos diários e sociais. Os comportamentos distintos direcionam a modelagem individual das residências, assim como foi realizado nos projetos. Dessa maneira, em consonância com o modo em que o *software* foi empregado em ambos projetos e com as conclusões relatadas na literatura, o procedimento de classificação foi individualizado. Em outras palavras, cada casa foi modelada separadamente, com suas respectivas séries padrão de consumo, e o algoritmo final foi definido como o que melhor representou a maioria.

Outro ponto importante apontado pelos pesquisadores foram as falhas cometidas pelo *software*. Houve citação de perfis de chuveiro que, por serem parecidos com a bacia sanitária, foram alocados como bacia por conta da hierarquia do *template*. Também foram apontadas classificações equivocadas entre bacias sanitárias e torneiras internas ou entre chuveiros e torneiras internas por apresentarem comportamentos muito próximos em alguns momentos. Houve eventos simultâneos que foram reconhecidos como um só equipamento.

### 4.3. CARACTERIZAÇÃO DOS DADOS

Os dados fornecidos pelos projetos têm dois formatos: os dados brutos (em .txt) que vieram diretamente dos *data loggers* e, na sucessão, os dados tratados pelo *Trace Wizard*® em conjunto com os pesquisadores, que fornecem os as séries temporais a nível individual (eventos simultâneos já separados).

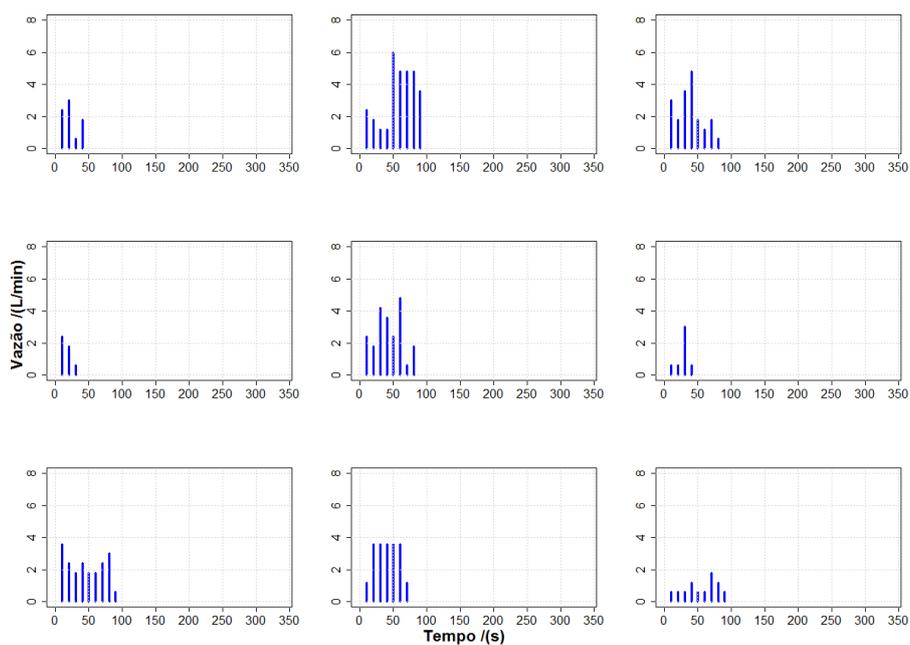
#### 4.3.1. PADRÕES – DADOS REFERENCIAIS

Como mencionado anteriormente, o monitoramento efetuado na pesquisa anterior compreendeu a fase investigativa e a partir dela foram recolhidas séries temporais de uso dos equipamentos para servirem de referência (padrões) de consumo. Nas Figuras 3, 4, 5 e 6 estão relacionados exemplos de perfis por equipamento de uma residência. Em análise inicial, percebe-se notável variação de padrões para um mesmo equipamento. Tal comportamento para torneira interna é subentendido (Figura 3), pois a mesma está associada a finalidades diversas, que variam em volume, duração e sequência das vazões. A torneira externa (Figura 4) em geral está associada a usos mais demorados e de alto consumo, por exemplo: lavagem de roupas, enchimento de baldes, limpeza de cômodos, entre outros. O uso do chuveiro (Figura 5) está principalmente associado a banho, mas varia de acordo com os hábitos de moradores (há a função de lavagem de roupas quando não existe torneira externa). A bacia sanitária em geral é caracterizada como aparelho automático, no entanto, o acionamento admite variação por parte do usuário (também observado por Nguyen et al. (2013a)), o que se reflete nos perfis observados (Figura 6). O percentual de perfis referência por equipamento e por residência pode ser visto na

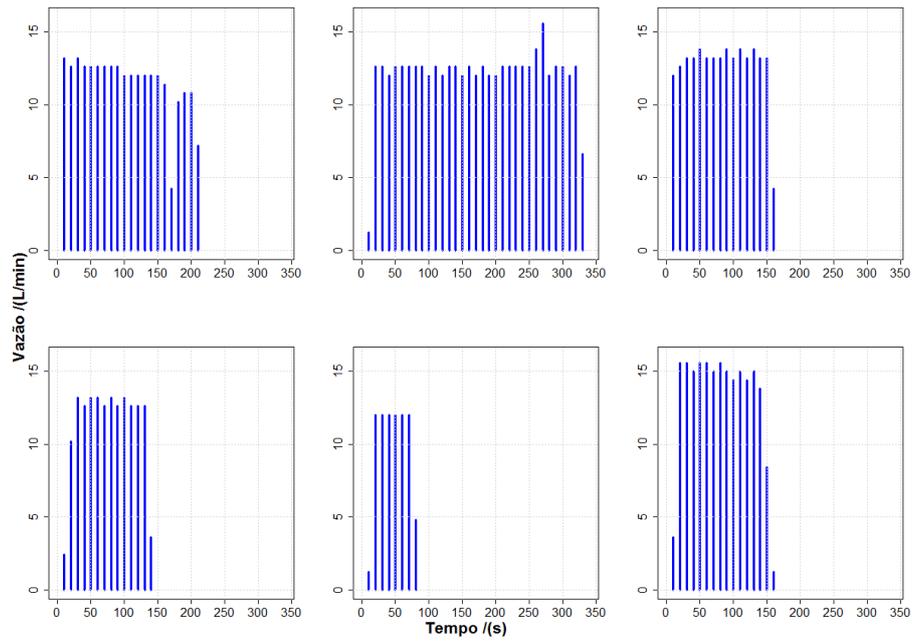
Tabela 2. Em geral, a torneira interna foi o equipamento mais acionado entre as residências.

*Tabela 2 - Distribuição percentual dos acionamentos por equipamento por residência no período investigativo.*

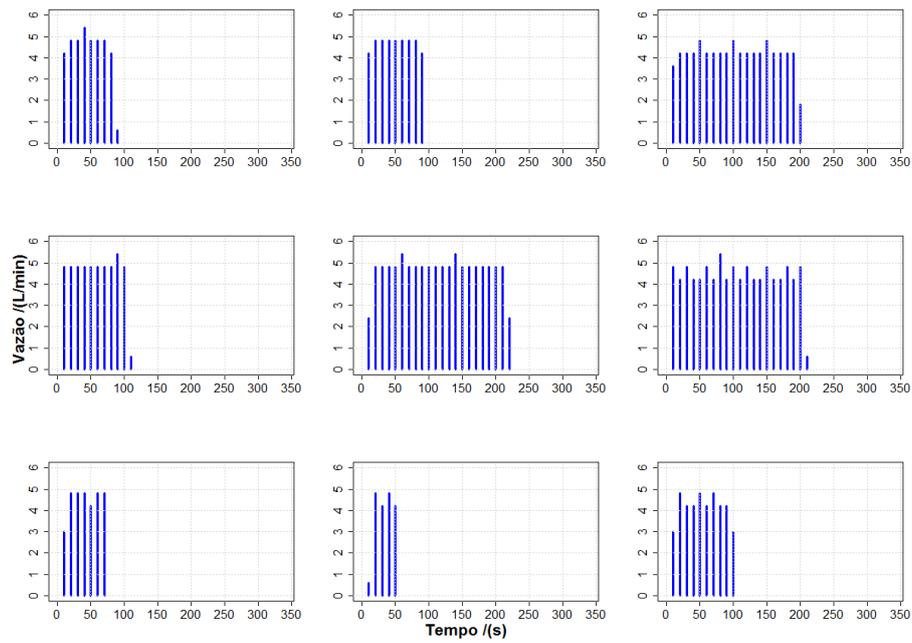
CASA	Bacia sanitária	Chuveiro	Torneira externa	Torneira interna
A	32%	17%	10%	41%
B	21%	33%	-	46%
C	19%	10%	1%	70%
D	15%	10%	-	75%
E	23%	17%	1%	59%
F	8%	38%	-	54%
G	15%	23%	-	62%
H	19%	14%	11%	56%
I	36%	29%	13%	22%
J	8%	11%	-	81%



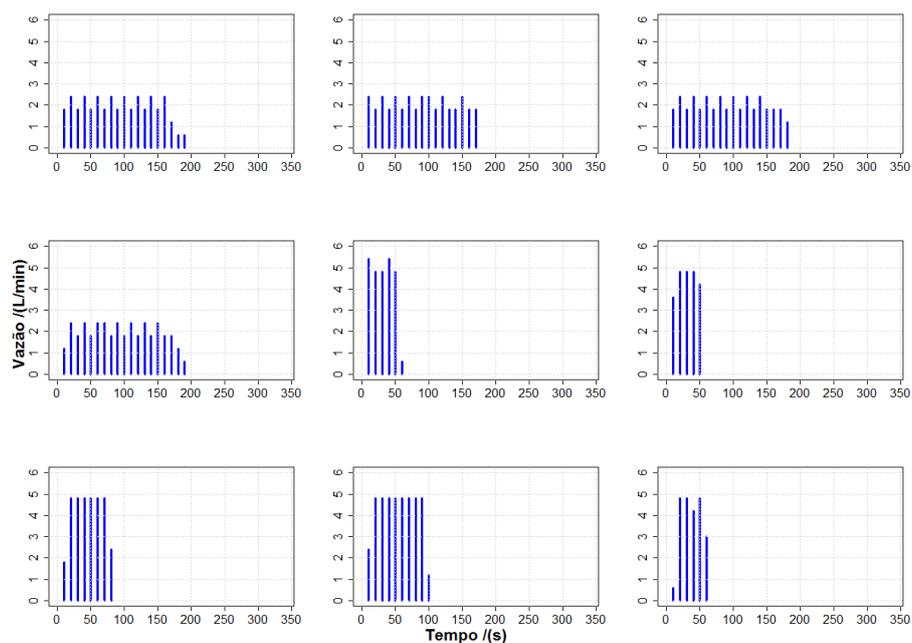
*Figura 3 - Exemplos de perfis de uso da torneira interna.*



**Figura 4** - Exemplos de perfis de uso da torneira externa.



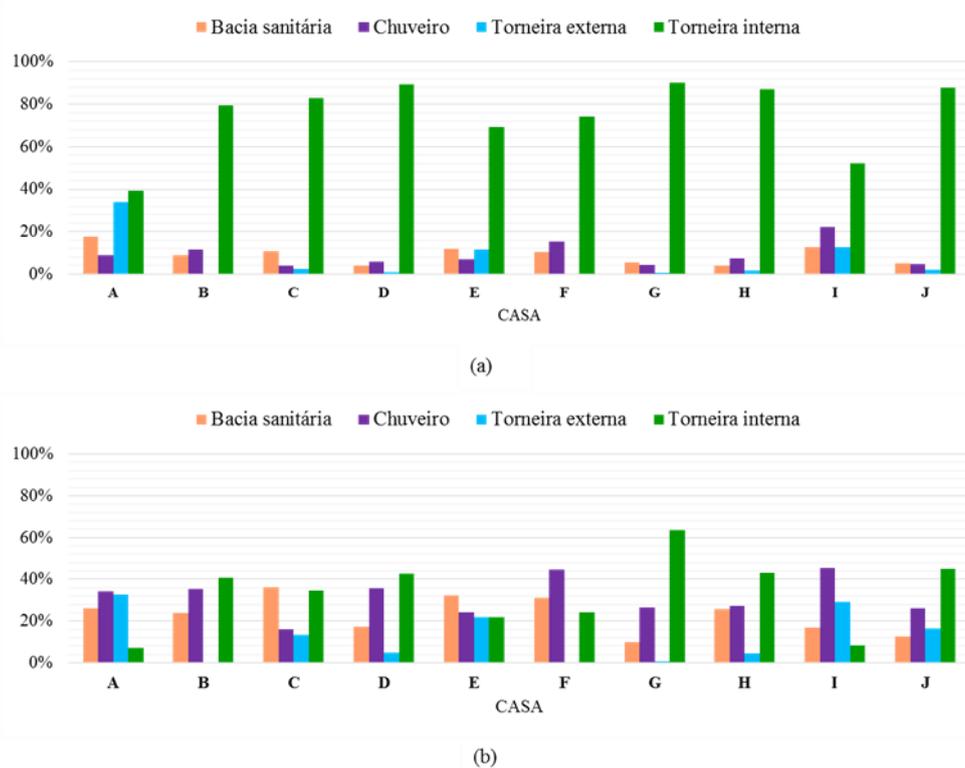
**Figura 5** - Exemplos de perfis de uso de chuveiro.



*Figura 6 - Exemplos de perfis de uso da bacia sanitária.*

#### 4.3.2. DADOS CLASSIFICADOS VIA TRACE WIZARD

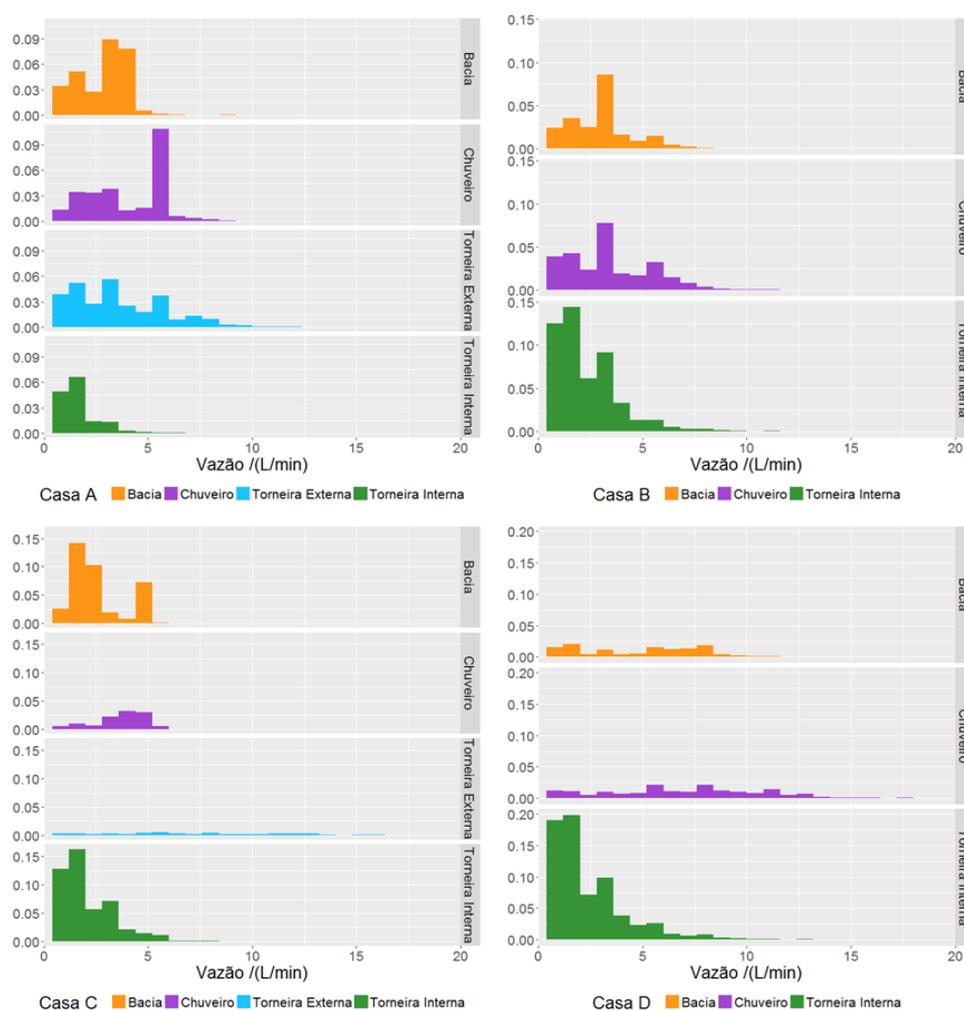
As torneiras internas foram as mais acionadas, em termos de recorrência de uso, em todas as residências; em termos volume as contribuições variam entre 7% e 45%, com exceção da casa G (64%) de onde foi sabido que o consumo desse equipamento foi bastante alto. Os outros aparelhos tiveram percentual de ocorrência inferior à torneira interna, entretanto contribuem significativamente em termos de volume (Figura 7).



**Figura 7** – Percentual de uso dos equipamentos em termos de quantidade de eventos (a) e de volume consumido (b), por casa.

Analisando as vazões por equipamento por domicílio através do uso de histogramas (Figuras 8, 9 e 10), percebeu-se que na maioria dos casos a ocorrência das vazões estava associada a mais de um equipamento. As intensidades de ocorrências das vazões por equipamento variaram, porém em geral constatou-se ser difícil presumir que um aparelho estava sendo acionado em detrimento de outro só com a informação da vazão. Os usos também podem ser caracterizados através da extração de características como duração, volume, moda e seu percentual de repetição, vazão máxima, média (FONTDECABA et al., 2013; NGUYEN et al., 2013a; VAŠAK et al., 2015). A duração (Figura 11) da torneira interna teve um range pequeno com relação aos outros, no entanto para durações superiores houve a presença de todos os outros equipamentos. Em termos de volume (Figura 12) em geral o intervalo entre o 1º e 3º quartis das torneiras internas foi estreito assim como da bacia sanitária; os chuveiros e as torneiras externas apresentaram um pouco mais de espalhamento. Como no caso das vazões, foi difícil determinar qual o equipamento estava sendo utilizado, somente com essas informações, porque uma característica que apresentava intervalo mais específico em um equipamento em uma casa, em outra já apresentava intervalo similar entre todos outros equipamentos. Os *boxplots* das características moda, vazão média e vazão máxima podem ser vistos no

Apêndice A (Figuras 28, 29 e 30), e assim como para o volume e a duração, suas informações não foram suficientes para classificar um uso. A exemplo de análise, na casa B, as distribuições de vazões para a bacia sanitária e chuveiro foram muito parecidas, logo a separação dessas tornou-se complicada, inclusive devido ao comportamento similar da duração e volume/ (Figuras 11 e 12). Já nas casas H e I, a duração da bacia sanitária foi mais específica; na casa J, o chuveiro apresentou bastante distinção entre volume, vazão média, vazão máxima e moda; na C a torneira externa apresentou moda, vazão média e vazão máxima com características um pouco distintas; esses pontos indicaram que a classificação desses equipamentos poderia ser eficiente.



*Figura 8 - Histogramas das vazões por aparelho, para as residências A, B, C e D.*

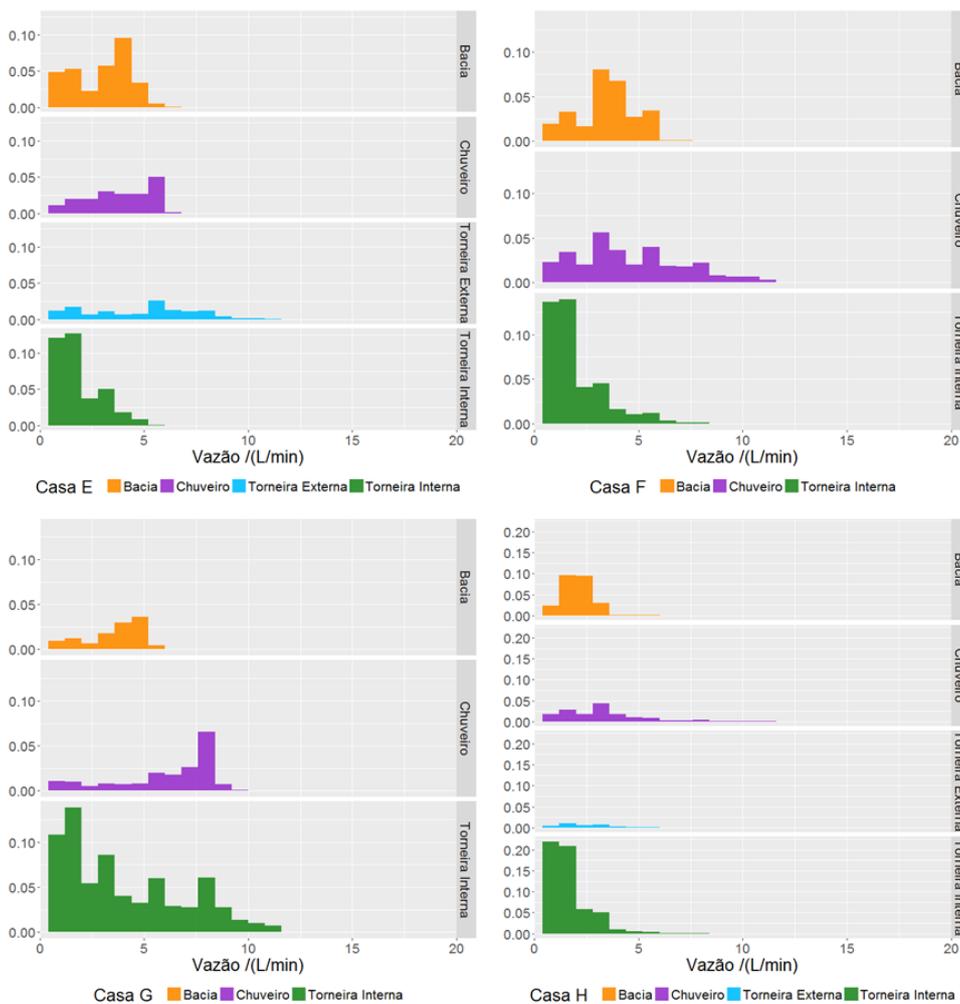


Figura 9 - Histogramas das vazões por aparelho, para as residências E, F, G e H.

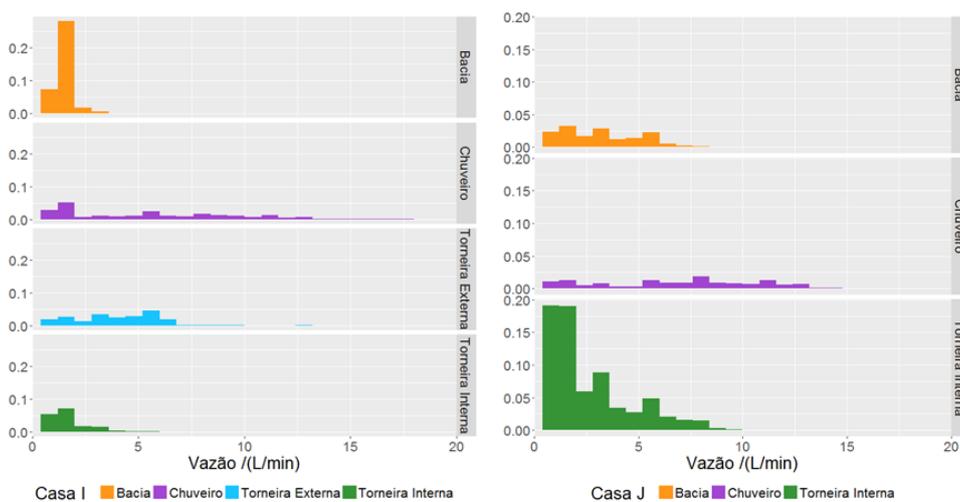
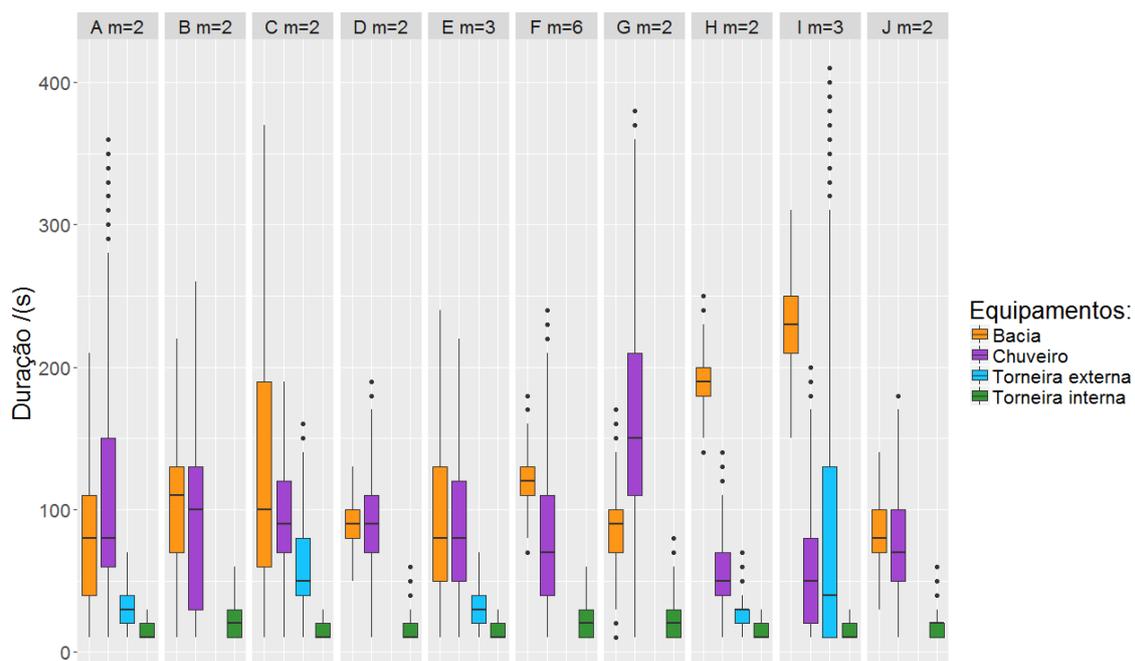
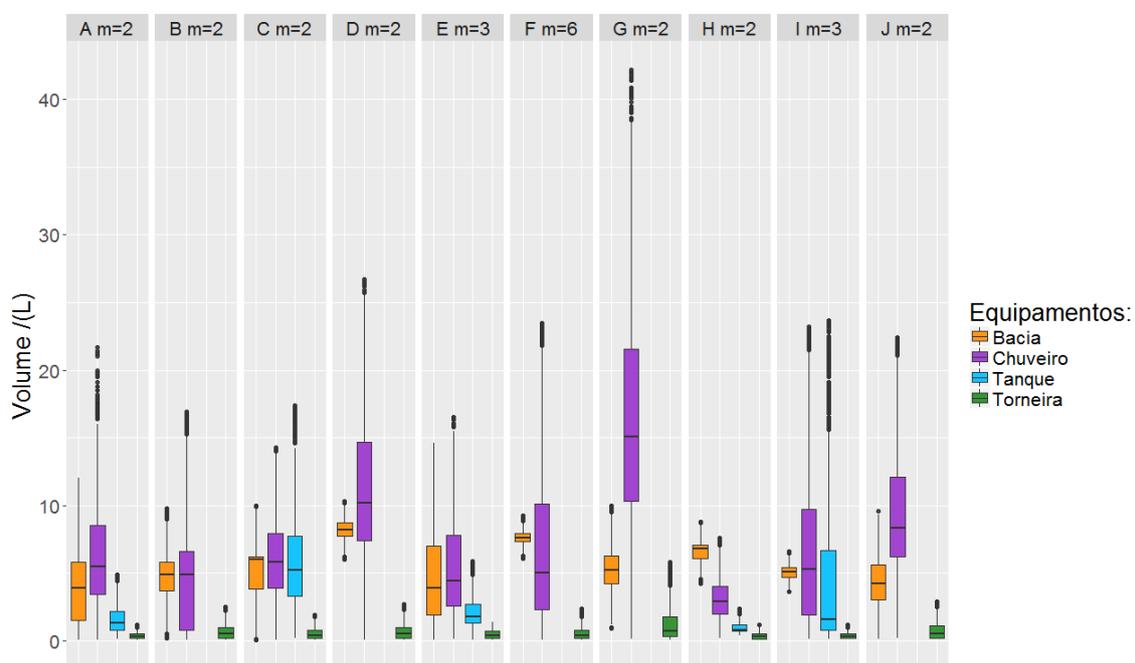


Figura 10 - Histogramas das vazões por aparelho, para as residências I e J.



*Figura 11 – Boxplot da duração dos equipamentos por casa.*



*Figura 12 - Boxplot do volume dos equipamentos por casa.*

## 5. MÉTODO

A ferramenta desenvolvida integra os algoritmos de separação de usos simultâneos e classificação de perfis. Dessa forma, são apresentados a seguir os métodos para cada um e ao final duas propostas de conexão desses algoritmos, uma vez que a ordem entre esses também foi alvo de estudo.

Vale ressaltar que a análise sobre o comportamento de consumo de cada casa foi estudada nos projetos anteriores, logo essa discussão não foi revisitada. Em adição, o modelo de classificação foi realizado a nível de equipamentos, sem subdivisões associadas aos tipos de uso por equipamento.

O algoritmo de identificação e separação de usos paralelos foi aplicado da mesma forma para todas as residências, ou seja, as considerações de comportamentos das séries temporais para compor as premissas foram construídas mediante análise conjunta de todas as residências.

### 5.1. IDENTIFICAÇÃO E SEPARAÇÃO DE USOS SIMULTÂNEOS

#### 5.1.1. FUNDAMENTAÇÃO

O estabelecimento do processo de separação demanda conhecimento sobre os comportamentos de consumo e determinação de premissas básicas. A noção de separação advém da percepção de interferência no processo de consumo, percebida pelo aumento da vazão em alguma seção do perfil base (NGUYEN et al., 2013b). Esse perfil base é uma referência adotada por apresentar consumo dominante e comportamento conhecido; tal interferência é percebida quando do aumento da vazão em seções aproximadamente constantes ou com tendência de queda. A princípio, o estabelecimento de regras para corte dos perfis não necessariamente produz resultados exatos, de acordo com as limitações citadas em seção anterior, ou seja, não é possível estabelecer exatamente onde a interferência começa e termina e nem quantos aparelhos estão envolvidos. No caso de equipamentos automáticos, a detecção de mudança no perfil é uma tarefa simples, pelo fato deles apresentarem comportamento pouco variável; já no caso de aparelhos manuais, o comportamento de uso de um equipamento não tem uma forma necessariamente rigorosa, o usuário é livre para variar a vazão no momento do uso e isso pode levar a falsa interpretação de usos múltiplos.

Foi efetuada intensa e detalhada avaliação empírica dos perfis que, em conjunto com análise das informações sobre o comportamento dos equipamentos, usou-se para melhor entender quais as características mais predominantes dos perfis e então estabelecer premissas que pudessem abranger a maior parte dos casos. Assim, foram adotadas as seguintes premissas:

- a) Não é possível verificar uso simultâneo num vetor de comprimento menor ou igual a 3, logo estes são considerados não simultâneos;
- b) A ocorrência de vazões repetidas são um indicativo ao perfil base, no entanto é necessário estabelecer uma porcentagem de ocorrência mínima (*pma*) que indique se o valor é de fato uma referência;
- c) O aumento da vazão quando a tendência é a estabilidade (moda) ou terminação (fim do uso) indica uso paralelo;

O algoritmo de corte estabelecido se baseou no uso do vetor gradiente, proposto por Nguyen et al. (2013b), com acréscimo do uso de estatísticas descritivas como moda e mediana para melhorar a qualidade da identificação. Assim como no trabalho mencionado, foi utilizado o parâmetro tolerância (*tol*) para indicar até que ponto vazões diferentes podem ser consideradas semelhantes. Dado um vetor  $X = (x_1, x_2, \dots, x_n)$  de comprimento  $N$ , seu gradiente  $G = (g_1, g_2, \dots, g_{n-1})$  é calculado através da Equação 1, que é uma adaptação levando em consideração a tolerância. Essa alteração auxilia no processo de varredura do perfil mostrando mais claramente as variações positivas. Por exemplo, se  $tol = 0,01l/s$ , as vazões  $0,05 l/s$  e  $0,06 l/s$  são consideradas parecidas e a diferença entre elas é considerada igual a zero,  $g_j = 0$ . Um paralelo entre o algoritmo desenvolvido e o proposto por Nguyen et al. (2013b) pode ser observado no Quadro 3.

$$g_j = \begin{cases} \frac{(x_j - x_{j-1})}{t_j - t_{j-1}}, & \text{se } (x_j - x_{j-1}) > tol \\ 0, & \text{se } (x_j - x_{j-1}) \leq tol \end{cases}, 2 \leq j < n \quad (1)$$

*Quadro 3 - Paralelo entre o algoritmo proposto e o de Nguyen et al., (2013b).*

<b>Tópicos</b>	<b>Função</b>	<b>Proposto por Nguyen et al. (2013b)</b>	<b>Proposto nesta pesquisa</b>
<i>Vetor gradiente</i>	Detectar a variação em vazões consecutivas	Foi aplicado numa versão modificada, em que há suavização no resultado.	O vetor gradiente adaptado de modo a considerar a tolerância (Equação 1)

<i>Tolerância</i>	Estabelece em que nível vazões são parecidas. Oferece flexibilização, que é importante por conta das variações de pressão na linha	Aplicada	Aplicada
<i>Estatística descritiva</i>	Apoiar o processo de tomada de decisão de corte	Não informação explícita sobre o uso da mesma	Aplicada
<i>Consideração sobre a primeira variação do gradiente</i>	Como a variação inicial está associada ao início da detecção, desconsidera-se o primeiro gradiente.	Aplicada	Aplicada

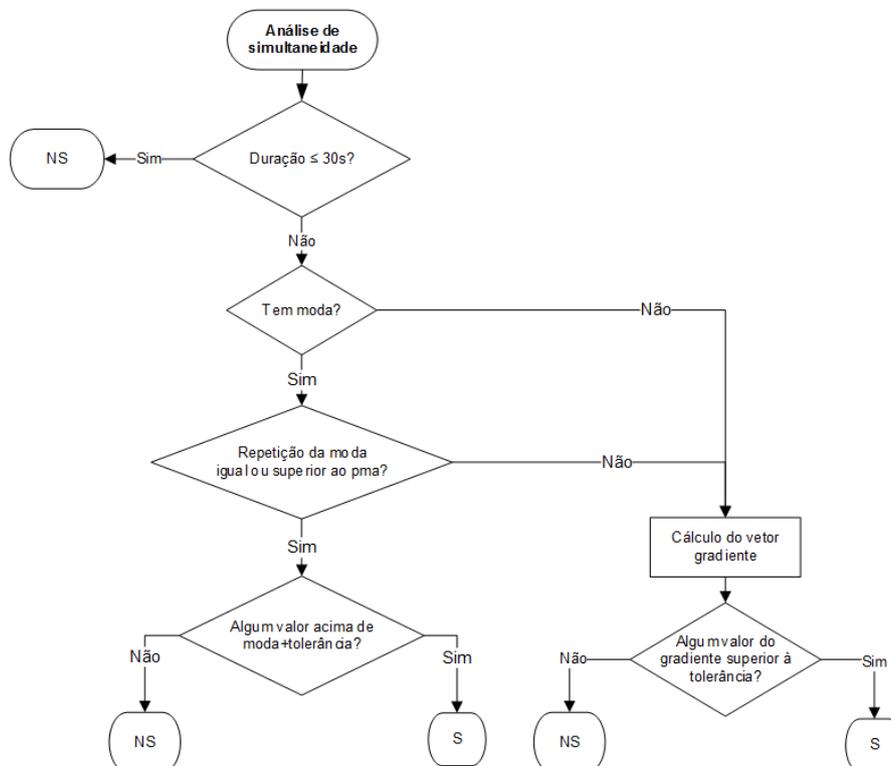
O algoritmo proposto é dividido em três seções: análise de simultaneidade, corte vertical e corte horizontal, a serem descritos nas seções seguintes.

### 5.1.2. CONSTRUÇÃO DO ALGORITMO

#### A. ANÁLISE DA SIMULTANEIDADE

Essa etapa é responsável pela análise inicial de indícios de ocorrência de simultaneidade e sua organização simplificada pode ser vista na Figura 13. A resposta que ela fornece é a indicação se o perfil é simultâneo e se sim, ele será submetido a posteriores processos de corte. A análise se inicia com a avaliação da duração e, de acordo com a premissa (**a**), perfis com duração menor ou igual a 30 segundos são considerados perfis não simultâneos (NS). Para durações superiores, a próxima análise é a ocorrência de moda, e sobre esse ponto é necessário considerar que podem haver flutuações na vazão por conta da variação de pressão na linha, produzindo séries com vazões alternadas ex.: 1212121. Em vista disso, é verificado se há mais de uma moda e se as mesmas são similares de acordo com a tolerância estabelecida: se parecidas, a moda será a média entre esses valores e o *pma* é calculado em função da frequência dessas vazões. Como dito na premissa (**b**), a repetição de determinada vazão leva a crer que ela pode representar o perfil base a depender do percentual *pma*. Se dado percentual calculado é maior ou igual ao *pma*, tal vazão passa a ser referência. Nesse caso, a próxima etapa é a verificação se há valores maiores que a moda: se existir, há indício de simultaneidade. Por outro lado,

se não houver moda é avaliado o vetor gradiente. O essencial é notar se há valores positivos de variação, em sendo assim, considera-se que há simultaneidade de usos.

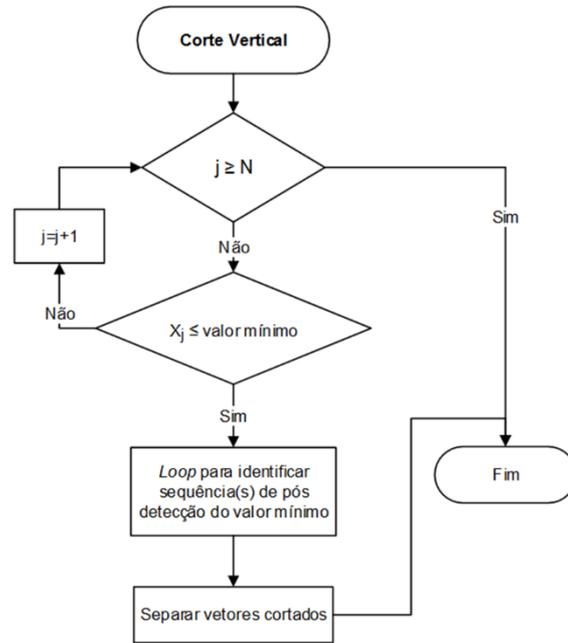


**Figura 13** - Fluxograma da etapa de análise de simultaneidade. S significa simultâneo e NS não simultâneo.

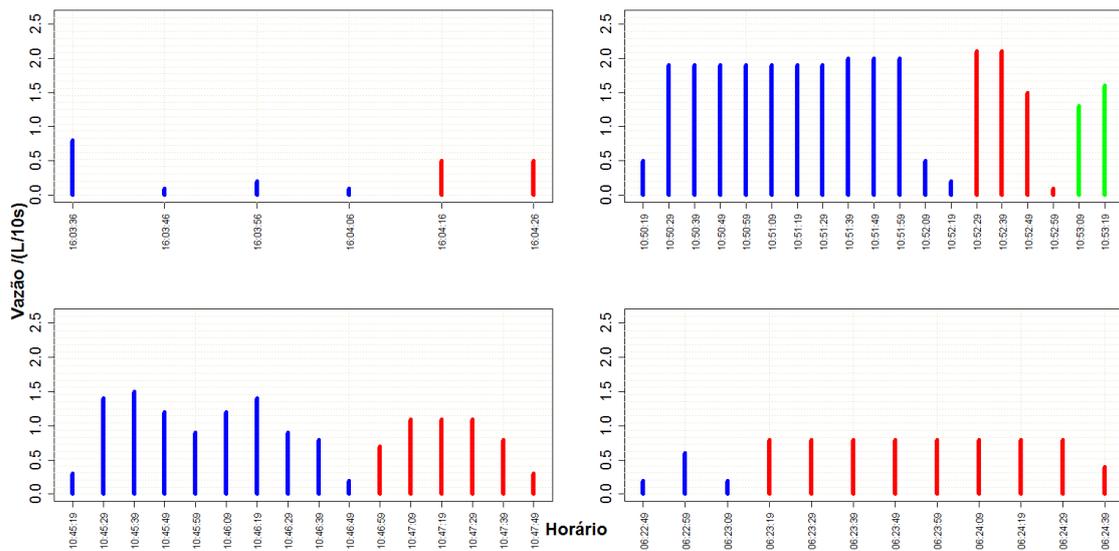
## B. CORTE VERTICAL

Como mencionado anteriormente, uma das grandes dificuldades dessa etapa é o desconhecimento do início e término de eventos paralelos nas séries. Por conta dessa limitação, o algoritmo de corte vertical foi estabelecido para situações em que ocorre aparente término do uso, mas que logo apresenta novo aumento. O valor mínimo considerado foi a resolução do medidor. A intenção é realizar uma varredura no perfil até encontrar o valor mínimo e partindo desse ponto continuar a leitura em busca de valores superiores a ele, de acordo com a tolerância. Aqueles que forem detectados são separados verticalmente do perfil original (Figura 14). Exemplos desse corte podem ser visualizados na Figura 15.

Os perfis que precisaram passar pelo corte vertical foram submetidos a outra análise de simultaneidade a fim de detectar se havia ainda perfis simultâneos. Se houverem, a próxima etapa é o corte horizontal.



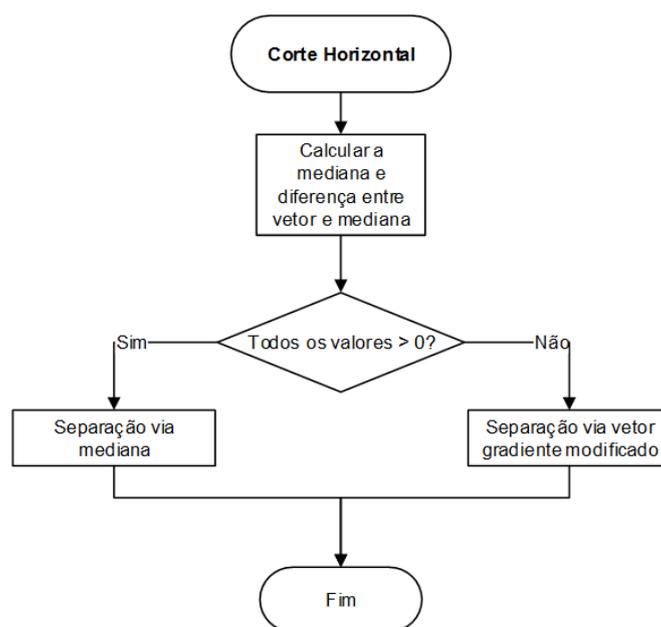
**Figura 14** - Fluxograma simplificado da etapa de corte vertical. A letra  $j$  representa a posição do vetor  $X$  de comprimento  $N$ .



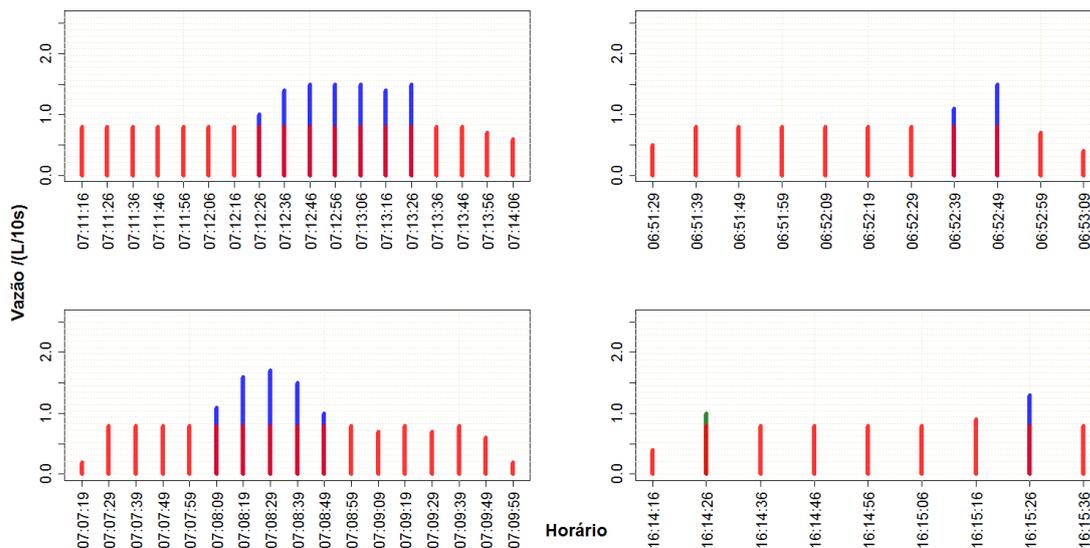
**Figura 15** - Exemplos de cortes verticais realizados. Há um perfil por gráfico (estado bruto) e a divisão por cores em cada um identifica como o algoritmo atuou.

### C. CORTE HORIZONTAL

O propósito dessa etapa é fazer separações baseadas na mediana ou no vetor gradiente modificado (Figura 16). Primeiramente é avaliada a diferença entre a série temporal e sua mediana. Se todos os valores forem positivos, o processo de corte se dá com a mediana como referência. Nesse contexto, a mediana é uma medida que auxiliou bastante na identificação de usos paralelos. De outra forma, o perfil é avaliado via varredura no vetor gradiente que busca seções em que há aumento de vazão. Quando o aumento é encontrado, o valor anterior ao aumento é adotado como referência e os perfis à frente que são superiores serão cortados ao nível dele. Por exemplo, considerando a sequência  $\{1, 2, 4, 5, 1\}$ , o seu gradiente correspondente é  $\{1, 2, 1, -4\}$ ; ainda se a  $tol = 1$ , o gradiente passa a ser  $\{0, 2, 0, -4\}$ . Na varredura do gradiente o primeiro valor maior que zero encontrado é 2 (referente as vazões  $\{2, 4\}$ ), logo a vazão 2 passa a ser uma referência e as vazões à frente são comparadas com ela, até que seja encontrado um valor menor ou igual, que nesse caso é 1. O corte então é aplicado e os perfis são separados da seguinte forma:  $\{1, 2, 2, 2, 1\}$  e  $\{2, 3\}$ ; ou seja, o valor referência passa a ficar no lugar das vazões superiores detectadas e as vazões superiores são subtraídas dessa referência e extraídas do perfil base. Exemplos dessa etapa podem ser visualizados na Figura 17. Vale ressaltar que após a identificação de um evento simultâneo, o processo de verificação continua até o final da série temporal seguindo as mesmas etapas.



**Figura 16** - Fluxograma simplificado da etapa de corte horizontal.



**Figura 17** - Exemplos de corte horizontal. Há um perfil (estado bruto) por gráfico e a divisão por cores em cada um identifica como o algoritmo atuou. O perfil base é o vermelho e em verde e azul estão as seções que foram cortadas.

### 5.1.3. PARÂMETROS

Para aplicar o algoritmo de separação foi necessário o estabelecimento dos valores de tolerância e percentual mínimo de aceitação da moda, ambos descritos anteriormente. A tolerância foi estabelecida baseando-se nas características da medição, ou seja, na resolução do medidor e no intervalo de medição, assim como também contou com a experiência do analisador com o comportamento dos dados. Dessa forma, o valor da tolerância foi variado entre 0l/s e 0,02l/s, em 0l/s têm-se tolerância zero entre vazões distintas; em 0,01l/s a tolerância é a resolução do medidor. Optou-se por não exceder 0,02l/s por conta da magnitude associada, afinal 0,03l/s equivale a 300ml lidos a cada dez segundos e entende-se que o mesmo era muito alto para considerar como tolerância. Os valores de *pma* variaram entre 40% e 100%. Para a avaliação inicial das etapas propostas, a tolerância inicial foi 0,01L/s e o percentual mínimo 50%.

### 5.1.4. ESTUDO DA CONFIGURAÇÃO DO ALGORITMO

Os processos de análise e corte sugeridos foram testados com base nos dados da semana investigativa. Os dados brutos (em .txt) passaram por um tratamento inicial que consistiu na organização das informações e conversão de unidades (pulso para vazão). Eles foram submetidos ao processo de separação, variando-se a aplicação das etapas de

acordo com o indicado no Quadro 4. O resultado dessa etapa indicou qual a configuração obteve o máximo acerto. O percentual de acerto foi calculado como a razão do *número de perfis em comum com identificados pelo grupo* e o *total de perfis identificados pelo grupo*. O passo seguinte foi a variação dos parâmetros de acordo com o citado no tópico anterior. Ao final, foi definido o formato de algoritmo mais adequado e o valor dos parâmetros que melhor se adequou para este específico estudo.

**Quadro 4** – Configurações testadas para determinar algoritmo final de identificação e separação de usos paralelos.

Configurações	Análise de simultaneidade	Corte Vertical	Corte Horizontal
1	X	X	-
2	X	X	X
3	X	-	X

## 5.2. CLASSIFICAÇÃO

### 5.2.1. FUNDAMENTAÇÃO TEÓRICA

#### 5.2.1.1. MEDIDAS DE (DIS)SIMILARIDADE

Uma medida de similaridade computa a força da relação entre dois objetos, ou seja, dois objetos serão similares se a medida estiver próxima do máximo valor possível. Uma medida de distância mede a dissimilaridade entre dois objetos e, de forma inversa à medida de similaridade, quanto maior o valor obtido menor o grau de parença entre os objetos. Uma distância poderá ser uma métrica se  $d: X \times X \rightarrow R$ , definida sobre um conjunto qualquer  $X$ , satisfaz as seguintes propriedades:

- $d(x, y) \geq 0$ ,
- $d(x, y) = 0$ , se e somente se  $x = y$ ,
- $d(x, y) = d(y, x)$ ,
- $d(x, z) \leq d(x, y) + d(y, z)$ ,  $\forall x, y, z \in X$ .

#### A. DTW

O DTW foi introduzido por Berndt e Clifford com o objetivo de permitir que uma série temporal fosse "esticada" para fornecer melhor combinação com outras séries temporais. Assim, dadas duas séries  $X = x_1, x_2, \dots, x_N$  e  $Y = y_1, y_2, \dots, y_M$ , o objetivo do algoritmo é encontrar um mapa que minimize a distância entre as séries,  $W$  (Equação 2)

com  $w_k = (i, j)_k$  que consiste no par de coordenadas  $x_i$  e  $y_j$ . Para alinhar as séries temporais, é construída a matriz de distorção  $D$ , de tamanho  $N$  versus  $M$ , cujos elementos são as distâncias  $d(x_i, y_j)$  (usualmente se utiliza a distância euclidiana) (KEOGH et al., 2005). Um esquema ilustrativo pode ser observado na Figura 18. A partir dela, de acordo com as restrições a seguir, o mapeamento é realizado:

$$W = w_1, w_2, \dots, w_K \quad (\max(M, N) \leq K < M + N - 1) \quad (2)$$

- **Condição de fronteira:**  $w_1 = (1, 1)$  e  $w_K = (N, M)$ , ou seja, começa nos primeiros pontos e termina últimos pontos das séries X e Y;
- **Condição monotônica:** Tanto os índices  $i$  como  $j$  permanecem os mesmos ou aumentam, eles nunca diminuem;
- **Continuidade (restrição local):**  $i(k) - i(k - 1) \leq 1$  e  $j(k) - j(k - 1) \leq 1$ , isto é, tanto  $i$  como  $j$  só podem aumentar em uma unidade em cada passo ao longo do caminho.

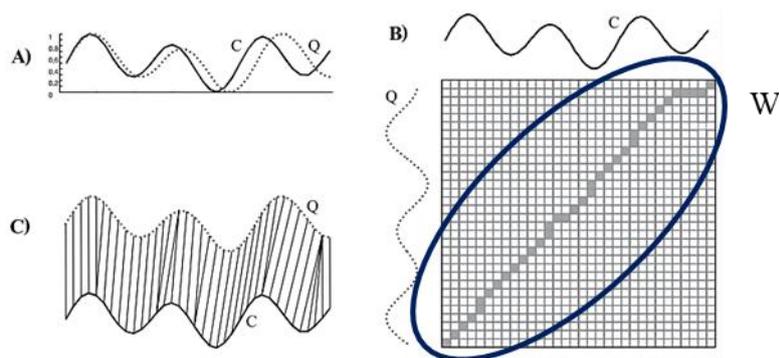
Como consequência das condições de continuidade e monotonicidade, tem-se a Equação 3.

$$w(k - 1) = \min \begin{cases} (i(k), j(k) - 1) \\ (i(k) - 1, j(k) - 1), \quad i, j \geq 1 \\ (i(k) - 1, j(k)) \end{cases} \quad (3)$$

A partir dessas regras, a  $D(N, M)$  mínima pode ser calculada pela Equação 4.

$$D(i, j) = \begin{cases} 0, & i = j = 0 \\ \infty, & i = 0, j > 0 \text{ ou } i > 0, j = 0 \\ d_{dtw}(x_i, y_j) + \min \begin{cases} D(i - 1, j - 1) \\ D(i - 1, j) \\ D(i, j - 1) \end{cases}, & i, j \geq 1 \end{cases} \quad (4)$$

$$d_{dtw}(x_i, y_j) = \begin{cases} |x_i - y_j|, & \text{se } x_i, y_j \text{ não são lacunas} \\ |x_i - y_{j-1}|, & \text{se } y_j \text{ é uma lacuna} \\ |y_j - x_{i-1}|, & \text{se } x_i \text{ é uma lacuna} \end{cases}$$



**Figura 18** - Esquema ilustrativo do alinhamento entre séries temporais, a partir do uso do DTW. Na letra A são exibidas as séries  $Q$  e  $C$ ; Na letra B é apresentada a matriz de distorção, com cada elemento sendo  $d(x_i, y_j)$  e ainda o caminho ( $W$ ) que oferece distância mínima entre as séries; e na letra C as séries são alinhadas de acordo com o caminho  $W$  (imagem de KEOGH et al., 2005).

Para maior aprofundamento sobre as características dessa medida de dissimilaridade, recomenda-se a leitura dos trabalhos Sakoe e Chiba (1978) e Keogh et al. (2005).

## B. EDR, ERP E LCSS

Os EDR (CHEN et al., 2005), ERP (CHEN et al., 2004) e LCSS (VLACHOS et al., 2002) são adaptações da distância de edição (do inglês *edit distance*) para séries numéricas. Essa vem do campo de comparação de séries de caracteres e computa o número de operações de inserção, exclusão ou substituição que são necessárias para fazer duas séries, de possíveis comprimentos diferentes, idênticas entre si. Para as séries numéricas, do mesmo modo que o DTW, constrói-se a matriz de distorção contendo  $d(x_i, y_j)$  e, de acordo com as regras de cada medida, calcula-se o quanto as séries são dis(similares), conforme o melhor alinhamento que pode haver entre elas.

O LCSS (Equação 5) computa similaridade entre séries  $X$  e  $Y$ , buscando a maior sequência em  $X$  que pode ser também encontrada em  $Y$ , ou seja, o resultado é o comprimento da maior subsequência comum; além disso não lida com operação de substituição. Nela há uma peculiaridade, a maior sequência não depende de termos consecutivos, basta ocorrerem na mesma ordem, em outras palavras, não considera lacunas. Segundo Kurbalija et al. (2014), a afirmação de que  $x_i = y_j$  é considerada forte para séries temporais, então esse termo passa a ser substituído por  $|x_i - y_j| \leq \varepsilon$  ( $\varepsilon$  é uma flexibilização análoga a variável *tol*). O EDR (Equação 7) procura o número mínimo de operações necessárias para transformar uma série na outra (medida de dissimilaridade) e

remove os efeitos de ruído ao quantificar a distância entre um par de elementos para dois valores possíveis: 0 e 1 (LCSS também realiza a mesma quantização); assim como o LCSS, faz uso do parâmetro  $\varepsilon$  para relativizar a diferença entre os valores da série. Já o ERP (Equação 6) cria uma constante  $g$  para quantificar as possíveis lacunas na série temporal e usa a distância L1 como o custo entre os elementos (medida de dissimilaridade). Os DTW, EDR e LCSS não seguem a desigualdade triangular e, portanto, não são métricas.

$$L(i, j) = \begin{cases} 0, & i = j = 0 \\ 1 + L(i - 1, j - 1), & i, j > 0 \text{ e } |x_i - y_j| \leq \varepsilon \\ \max(L(i - 1, j), L(i, j - 1)), & i, j > 0 \text{ e } |x_i - y_j| > \varepsilon \end{cases} \quad (5)$$

$$E(i, j) = \begin{cases} \sum_{k=1}^j |y_k - g|, & i = 0 \\ \sum_{k=1}^i |x_k - g|, & j = 0 \\ \min \begin{cases} d_{erp}(x_i - y_j) + E(i - 1, j - 1) \\ |x_i - g| + E(i - 1, j) \\ |y_j - g| + E(i, j - 1) \end{cases}, & i \neq 0 \text{ e } j \neq 0 \end{cases} \quad (6)$$

$$d_{erp}(x_i, y_j) = \begin{cases} |x_i - y_j|, & \text{se } x_i, y_j \text{ não são lacunas} \\ |x_i - g|, & \text{se } y_j \text{ é uma lacuna} \\ |y_j - g|, & \text{se } x_i \text{ é uma lacuna} \end{cases}$$

$$E(i, j) = \begin{cases} j, & i = 0 \\ i, & j = 0 \\ \min \begin{cases} d_{edr}(x_i, y_j) + E(i - 1, j - 1) \\ 1 + E(i - 1, j) \\ 1 + E(i, j - 1) \end{cases}, & i \neq 0 \text{ e } j \neq 0 \end{cases} \quad (7)$$

$$d_{edr}(x_i, y_j) = \begin{cases} 0, & |x_i - y_j| \leq \varepsilon \\ 1, & |x_i - y_j| > \varepsilon \end{cases}$$

No Quadro 5 estão resumidas as informações importantes sobre as medidas de (dis)similaridade estudadas.

*Quadro 5 - Resumo de informações sobre as medidas de (dis)similaridade.*

Distâncias	DTW	LCSS	ERP	EDR
Desigualdade triangular	Não satisfaz	Não satisfaz	<b>Satisfaz</b>	Não satisfaz

<b>Valor que identifica séries iguais</b>	0	Comprimento do menor vetor	0	0
<b>Lacuna</b>	Considera	Não considera	Considera	Não considera
<b>Ruído</b>	Sensível	Não sensível	Sensível	Não sensível
<b>Uso de parâmetros</b>	Não	$\epsilon$	$g$	$\epsilon$
<b>Tempo computacional</b>	$O(nm)$ . O tempo é superior a euclidiana, $O(n)$ , no entanto, as séries temporais são curtas, sendo o tempo mais afetado pelo número de amostras e padrões. (n e m são os comprimentos das series temporais sob comparação)			

### C. OTIMIZAÇÃO DE PARÂMETROS

Em termos de otimização do DTW, é possível alterar a restrição local e as variações mais aplicadas foram desenvolvidas por Sakoe e Chiba (1978). Eles estabeleceram o parâmetro  $P$  (Equação 8) que representa uma restrição de inclinação, com  $m$  sendo o número de passos na direção da abcissa ( $x$ ) e  $n$  é o número na direção da ordenada ( $y$ ). A intenção é restringir os movimentos para prevenir alinhamentos entre sequências com comprimentos muito diferentes. Se  $P = 0$  não há restrições na inclinação da função de distorção; se  $P = \infty$ , o caminho é restrito à diagonal.

$$P = \frac{n}{m} \quad (8)$$

Além desse, é aplicada uma função de peso  $s(k)$  cuja função é normalizar  $D(X, Y)$  (Equação 11) e dar capacidade de flexibilização com duas alternativas, simétrica e assimétrica. Elas são aplicadas na no formato das equações 9 e 10.

$$s(k) = (i(k) - i(k - 1)) + (j(k) - j(k - 1)) \rightarrow N = I + J \quad (9)$$

$$s(k) = (i(k) - i(k - 1)) \rightarrow N = I \quad (10)$$

$$D(X, Y) = \min_W \left[ \frac{\sum_{k=1}^K d(w(k)) \cdot s(k)}{\sum_{k=1}^K s(k)} \right] = \frac{1}{N} \min_W \left[ \sum_{k=1}^K d(w(k)) \cdot s(k) \right] \quad (11)$$

Em termos da Equação 4, a Equação 11 entra na seção em que  $i \neq 0$  e  $j \neq 0$ , substituindo o cálculo do mínimo anterior aos  $x_i$  e  $y_j$  (Equação 12).

$$\min \begin{cases} D(i-1, j-1) + s(k).d(x_i, y_j) \\ D(i-1, j) + s(k).d(x_i, y_j) \\ D(i, j-1) + s(k).d(x_i, y_j) \end{cases} \quad (12)$$

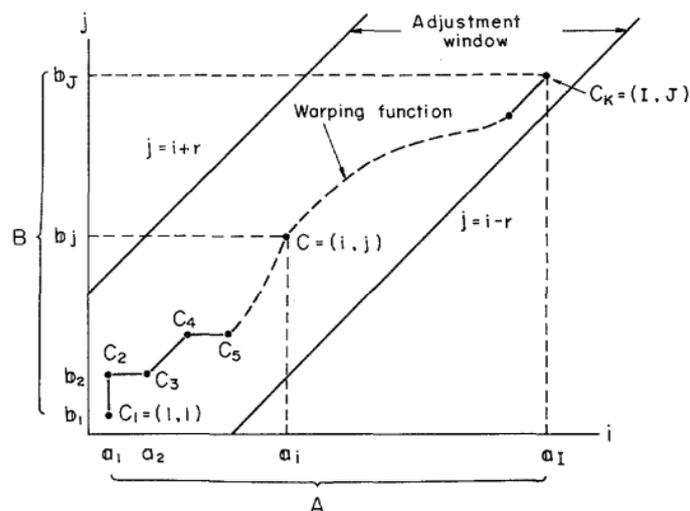
No caso da simétrica o peso leva em conta os passos de  $i$  e  $j$ , se há um passo apenas para  $i-1$  ou  $j-1$ ,  $s(k) = 1$ ; se há movimento para  $i-1$  e  $j-1$ ,  $s(k) = 2$  (Equação 13). Para a assimétrica,  $s(k)$  só depende de passos em  $i$  (Equação 14).

As variações utilizadas foram: o modo padrão (Equação 4) e as simétrica e assimétrica com  $P = 0$  (Equações 13 e 14). Outros valores de  $P$ , com e sem simetria, foram testados, mas devido ao grande intervalo de comprimentos das séries temporais, elas não funcionam.

<b>Simetria</b>	<b>P</b>		
Simétrico	0	$\min \begin{cases} D(i-1, j-1) + 2 * d_{dtw}(x_i, y_j) \\ D(i-1, j) + d_{dtw}(x_i, y_j) \\ D(i, j-1) + d_{dtw}(x_i, y_j) \end{cases}$	(13)
Assimétrico	0	$\min \begin{cases} D(i-1, j-1) + d_{dtw}(x_i, y_j) \\ D(i-1, j) + d_{dtw}(x_i, y_j) \\ D(i, j-1) + 0 * d_{dtw}(x_i, y_j) \end{cases}$	(14)

Para os LCSS e EDR os valores testados para  $\varepsilon$  são: 0l/s, 0,01l/s e 0,02l/s. O parâmetro  $g$  para o ERP não foi modificado, já que é uma referência didática, relacionada com o sistema cartesiano.

Existe a possibilidade de impor restrições globais, que partindo do princípio de que um caminho longe da diagonal é provavelmente ruim, tencionam restringir de acordo com uma janela de comprimento  $r$  (Figura 19).

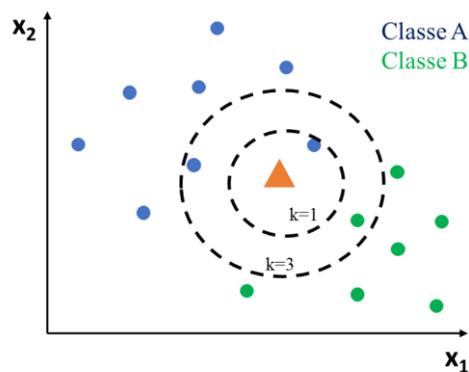


**Figura 19** – Esquema ilustrativo da restrição da janela por Sakoe e Chiba (1978).

Essas restrições podem ser aplicadas a todas as medidas de (dis)similaridade avaliadas aqui, no entanto o tamanho da janela é função do comprimento das séries. Como há grande variedade de comprimentos inviabiliza-se essa aplicação, pois não há limite para o tamanho das séries temporais.

#### 5.2.1.2. ADAPTAÇÃO DO ALGORITMO K-NEAREST NEIGHBORS

Para avaliar a eficácia de cada medida foi construído um modelo de classificação baseado no conceito do kNN. Este algoritmo de aprendizado é não paramétrico e, sendo do tipo *lazy-learning* (aprendizagem preguiçosa), não há parâmetros a serem treinados. Na classificação k-NN, o resultado é a atribuição de um objeto à classe mais comum dentre seus  $k$  vizinhos mais próximos, com  $k$  sendo um número natural maior que zero (WONDERS et al., 2016). Se  $k = 1$ , o objeto é atribuído à classe do vizinho mais próximo (Figura 20). A distância entre os objetos é o cálculo mais importante do processo e é realizado via uso de medida de (dis)similaridade.



**Figura 20** - Imagem ilustrativa do emprego do  $k$ -NN. O triângulo laranja é o objeto a ser classificado. Para  $k=1$ , ele corresponde a classe A. Para  $k=3$ , ele também corresponde a classe A, por essa ser a classe dominante (2 de A versus 1 de B).

Essa técnica lida com vetores num espaço  $n$ -dimensional e não com séries temporais, logo foi proposto um algoritmo adaptado que compare um perfil teste com os perfis padrão (exemplo ilustrativo na Figura 18) e que seja selecionada a classe de acordo com o mais próximo, o que equivale a  $k = 1$ .

### 5.2.2. MODELOS PROPOSTOS

Para analisar a viabilidade do uso das medidas de similaridade, aproximadamente 20% dos dados foram dispostos para testes. Embora o percentual aparente ser pequeno, ele equivale a aproximadamente 60.000 perfis de consumo e por conta das variações propostas a seguir, esse valor foi determinado para reduzir o tempo computacional.

Dois modelos foram propostos: modelo sem restrição (MSR) em que cada perfil a ser classificado é comparado com padrões de todos os equipamentos e modelo com restrição (MCR), no qual são calculadas a duração e o volume do perfil para pré-selecionar quais equipamentos listados seguem para a classificação.

Os critérios de volume e duração foram escolhidos na intenção direcionar ao classificador equipamentos que abrangessem comportamento similar a série temporal em teste, auxiliando em uma determinação mais assertiva. Nesse caso, para a seleção do equipamento, são comparados os volume e duração do perfil teste com a média  $\pm$  desvio de cada equipamento: se pelo menos um dos critérios passar, o equipamento é escolhido. A estrutura dos modelos pode ser vista na Figura 21.

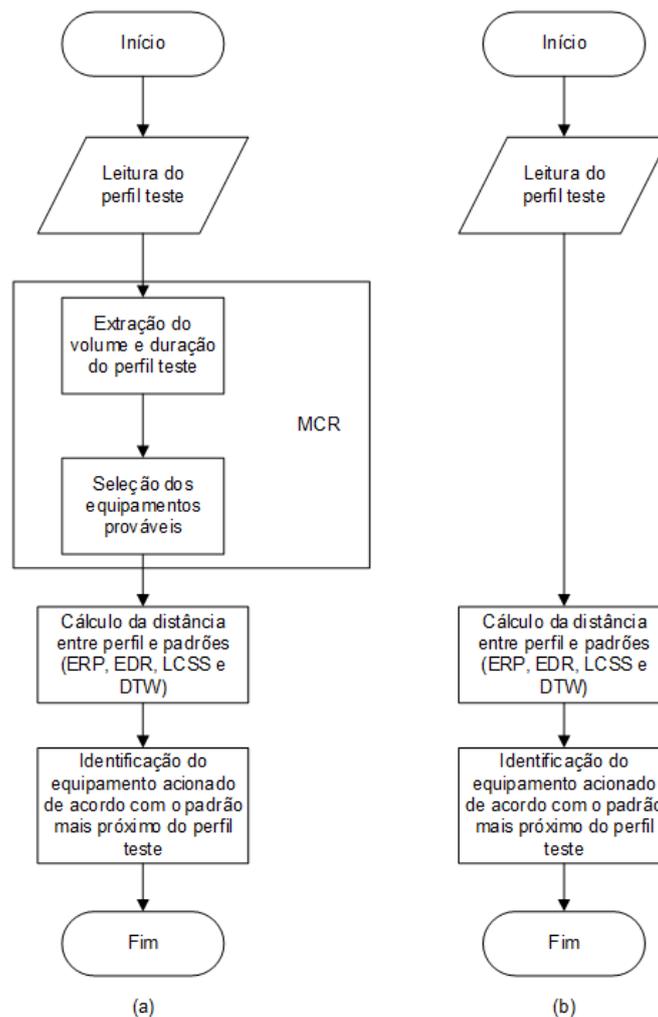


Figura 21 – Fluxograma simplificado dos modelos MCR (a) e MSR (b)

Quanto às medidas de (dis)similaridade os parâmetros selecionados foram:

- *Threshold* –  $\varepsilon$ : Esse parâmetro, descrito anteriormente, está relacionado com as distâncias **LCSS** e **EDR**. Para ambas houve variação com os seguintes valores: 0l/s; 0,01l/s e 0,02l/s.
- *Penalização de lacunas* –  $g$ : Este está relacionado com a distância **ERP** e não foi modificado, já que é uma referência didática, relacionada com o sistema cartesiano.
- *Restrições locais*: Associadas à distância DTW, as opções utilizadas foram as variações SP0 (simétrica com  $P = 0$ ) e ASP0 (assimétrica com  $P = 0$ ), além da básica.

Os modelos e as conseqüentes respostas de cada medida de (dis)similaridade foram avaliados através de três indicadores de qualidade, a precisão, sensibilidade e

acurácia. A precisão (Equação 15) está relacionada com a qualidade do classificador, indicando qual o percentual dos perfis que foram assinalados para a classe correta. Já a sensibilidade (Equação 16) indica o percentual dos dados originais de determinada classe que foram classificados corretamente. Ambos serão avaliados tanto em quantidade de eventos quanto em volume. Os critérios de avaliação devem priorizar a medida de (dis)similaridade que consiga detectar a maior quantidade de volume consumido (alta sensibilidade), visto que a intenção futura é poder derivar informações, que serão pertinentes se caracterizarem a maior parte do consumo. Na sequência, é importante que o classificador detecte corretamente os equipamentos acionados (alta precisão), assim serão avaliados os resultados tanto em termo de quantidade de eventos quanto em volume reconhecido. A acurácia (Equação 17), que indica o percentual de acerto agregando todas as classes, será utilizada na avaliação inicial dos resultados globais. Na Figura 22 é exibido uma ilustração da matriz de confusão para mais de uma classe.

		Real		
		Classe A	Classe B	Classe C
Predito	Classe A	Verdadeiro A	Falso A	Falso A
	Classe B	Falso B	Verdadeiro B	Falso B
	Classe C	Falso C	Falso C	Verdadeiro C

**Figura 22** –Matriz de confusão ilustrativa, que situa as informações utilizadas para cálculo dos indicadores. A exemplo, nessa estão os nove resultados possíveis em uma tabela de contingência  $3 \times 3$ .

$$Precisão = \frac{Verdadeiro X}{Preditos X} \quad (15)$$

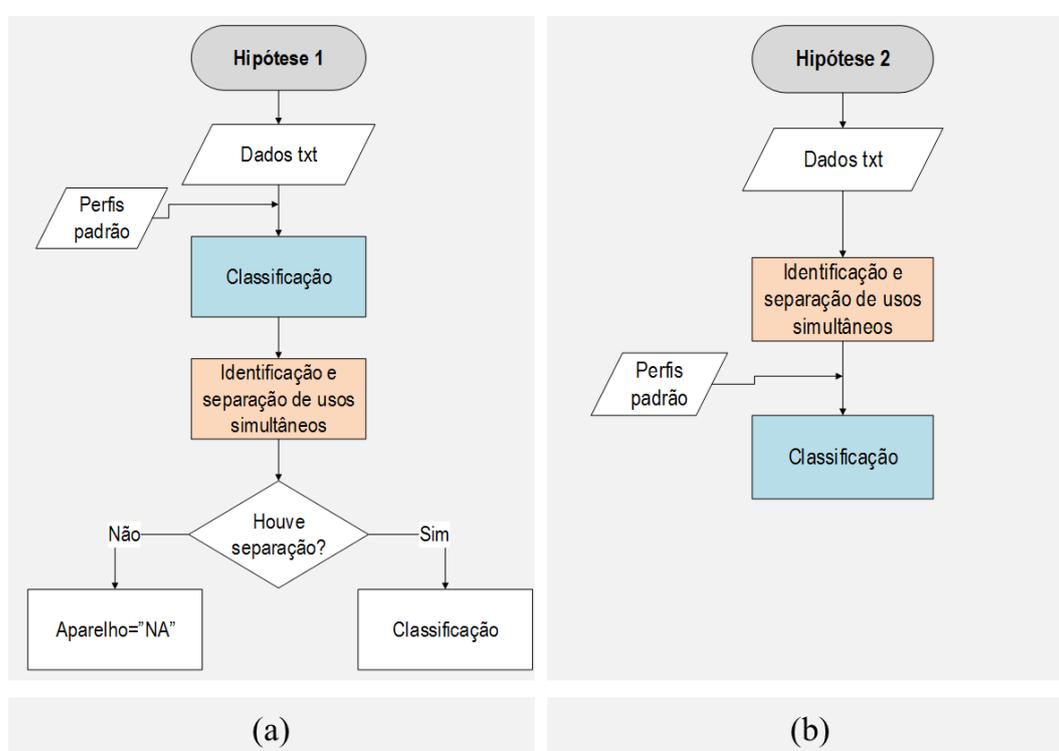
$$Sensibilidade = \frac{Verdadeiro X}{Real X} \quad (16)$$

$$Acurácia = \frac{\sum Verdadeiro}{Total} \quad (17)$$

Todos os algoritmos foram programados via uso do *software* R. Em termos de tempo computacional, o uso de medidas de (dis)similaridade (lembrando que são algoritmos) envolvem mais tempo para computar resultados quando comparados com a distância euclidiana, que é um cálculo direto. No entanto, as séries temporais em geral têm comprimentos curtos então o processo de cálculo não foi longo. Assim o impacto estava associado ao número de séries testadas e o número de padrões no banco.

### 5.3. ALGORITMO INTEGRADO

A proposta foi a integração dos algoritmos de classificação e separação de usos simultâneos, para tanto foram propostas duas hipóteses (Figura 23). Na hipótese 1, parte-se do princípio que após a classificação, se o resultado indicou múltiplas respostas possivelmente há simultaneidade de usos. Para esses a avaliação de simultaneidade é aplicada e caso não haja detecção de usos simultâneos, é redundante tentar reclassificar e assim se atribui classe nula. Na hipótese 2 os perfis inicialmente são submetidos a etapa de separação de usos paralelos e posterior classificação. Para testar, foram escolhidos, aleatoriamente, dados da casa C.



**Figura 23** - Fluxograma da hipótese 1 (a) iniciando com a classificação e da hipótese 2 (b) iniciando com a análise e separação de perfis simultâneos.

## 6. RESULTADOS

### 6.1. IDENTIFICAÇÃO E SEPARAÇÃO DE USOS SIMULTÂNEOS

Ao todo na semana investigativa foram registrados 9059 perfis no estado bruto, ou seja, sem nenhum processo de separação. Após análise do grupo de pesquisa, houve a identificação de 10135 perfis, demonstrando que ocorreram casos de uso simultâneo de equipamentos. Os resultados das configurações testadas estão registrados na Tabela 3. Na residência F houveram problemas com os dados .txt e por isso não entraram nessa etapa. A princípio, foram comparados os perfis brutos com os já tratados no intuito de identificar quantos deles foram considerados não simultâneos. Como resultado, 76,4% dos dados brutos foram perfis cujos acionamentos foram individualizados. A configuração 1 indicou 76% de perfis em comum. Quando, na sequência do corte vertical, os dados foram submetidos ao corte horizontal, ocorreu um aumento no percentual de acerto (79%). De outra forma, a configuração 3 mostrou melhor potencial de acerto, 81%, demonstrando que a seção de corte vertical demanda definição de outros critérios além do utilizado. Esse resultado, embora generalizado, se repetiu em todas as residências.

*Tabela 3 - Percentual médio de acerto das configurações sugeridas.*

Configurações	Análise de simultaneidade	Corte Vertical	Corte Horizontal	Percentual médio de acerto
1	X	X	-	76%
2	X	X	X	79%
3	X	-	X	81%

Na avaliação da mudança dos parâmetros (Tabela 4), os resultados indicaram que a tolerância de 0,011/s foi aquela que produziu os melhores resultados. A variação dos percentuais de acerto com a variação do *pma* foi pequena, mas identificou-se que o mais indicado foi o valor de 90% (para maior abrangência em outras aplicações). Ou seja, de acordo com o modo em que as séries temporais foram avaliadas previamente pelo grupo, o algoritmo entra em consonância quando admite que a moda é significativa quando predomina em pelo menos 90% do total. Essas conclusões são iguais em todas as residências em teste. Em suma, o algoritmo de separação indicado foi aquele que reuniu as etapas de análise de simultaneidade mais o corte horizontal e os parâmetros adequados foram  $tol = 0,011/s$  e  $pma = 90\%$ .

*Tabela 4 - Resultado médio do percentual de acerto com as variações de parâmetros.*

<i>pma</i>	Tolerância (l/s)		
	0	0,01	0,02
<b>40%</b>	73,2%	80,8%	78,3%
<b>50%</b>	73,2%	81,1%	78,5%
<b>60%</b>	73,2%	81,5%	78,8%
<b>70%</b>	73,6%	81,7%	79,0%
<b>80%</b>	73,7%	81,7%	79,0%
<b>90%</b>	73,8%	81,8%	79,0%
<b>100%</b>	73,8%	81,8%	79,0%

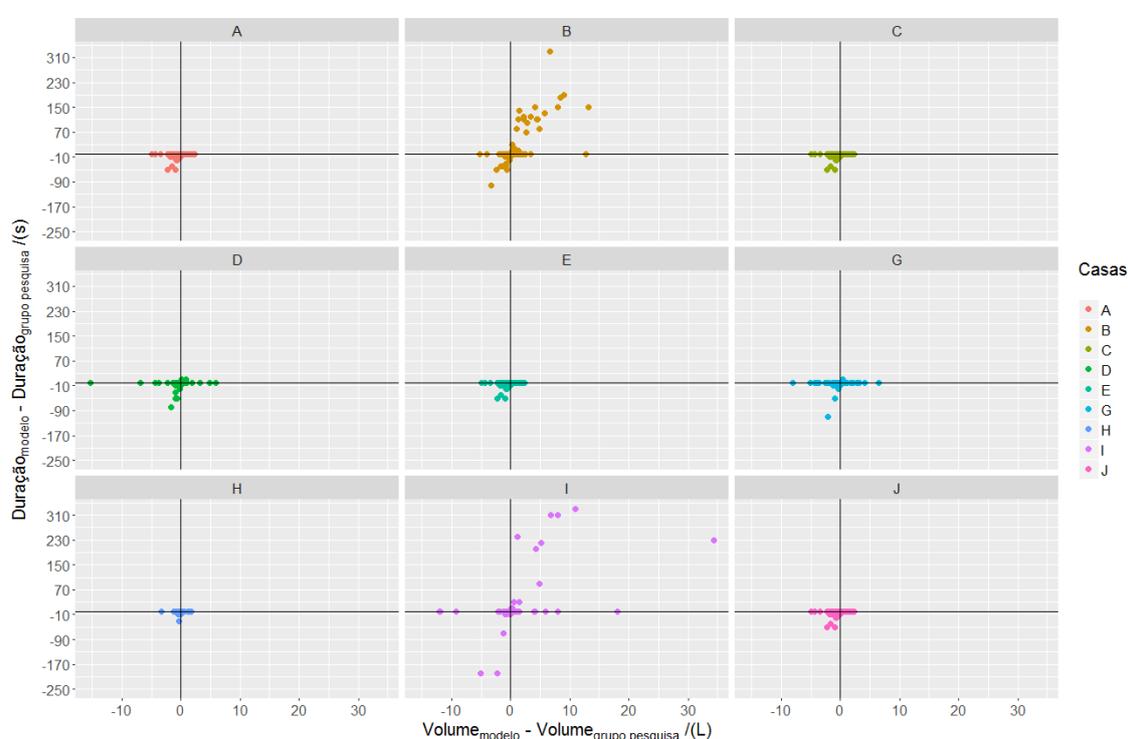
Com a definição dos valores adequados dos parâmetros, analisou-se as discrepâncias entre os resultados fornecidos pelo algoritmo e os dados previamente tratados (Tabela 5). Em geral, as separações promovidas pelo algoritmo variaram entre 68% e 94% de acerto. Os erros foram qualificados entre: perfis que não tinham compatibilidade em horário inicial ou com volume e/ou duração distintos. No primeiro caso, esse erro decorreu de cortes horizontais em horários distintos ao que foram considerados, ou ainda da ausência de séries temporais (no banco final de padrões) por conta da dificuldade na classificação ou vazamentos. No segundo caso, os horários foram compatíveis, mas assim como no anterior, houveram considerações que levaram a cortes distintos e esses erros foram melhor detalhados na Figura 24. Eles se dividiram entre o primeiro e terceiro quadrante do gráfico. No primeiro quadrante estão os casos em que os volume e duração pelo algoritmo são superiores aos dados referência, isso mostra que houve realização de cortes verticais por parte do grupo (no algoritmo vigente não há esta parte). No terceiro quadrante ocorreu o inverso, os volume e duração pelo algoritmo foram inferiores aos dados referência; isso denotou que os cortes horizontais foram realizados com critérios distintos, em que o algoritmo realizou um corte com menor duração e por consequência menor volume. Nas residências A, C, D, E, G, H e J há predomínio dos dados no terceiro quadrante e vários casos de perfis com durações iguais mais volumes diferentes. Nas casas B e I aconteceram mais erros no primeiro quadrante que as outras, evidenciando que nelas houve mais episódios de cortes verticais.

Vale ressaltar que o algoritmo desenvolvido oferece flexibilidade para escolha não só de parâmetros, como também das etapas. Esse atributo é importante, uma vez que outros trabalhos que venham a fazer uso dele podem empregar outros tipos de medidores,

variando resolução e intervalo de medição, assim como ter critérios de corte que só admitam algum dos tipos de cortes ou todos.

**Tabela 5** - Resultados do processo de separação, com informação do percentual de acerto e de erros, separados em duas categorias: incompatibilidade horária e diferenças entre duração ou volume.

		A	B	C	D	E	G	H	I	J
<b>Percentual de acerto</b>		84%	87%	86%	72%	82%	88%	94%	90%	68%
<b>Erros</b>	<b>Erros em duração e /ou volume</b>	14%	9%	11%	11%	15%	10%	4%	8%	22%
	<b>Incompatibilidade do horário inicial</b>	2%	3%	2%	17%	4%	2%	2%	2%	9%



**Figura 24** - Gráfico de dispersão da diferença entre o volume do perfil pelo algoritmo e pelo grupo (eixo x) versus a diferença entre a duração do perfil pelo algoritmo e pelo grupo, por residência.

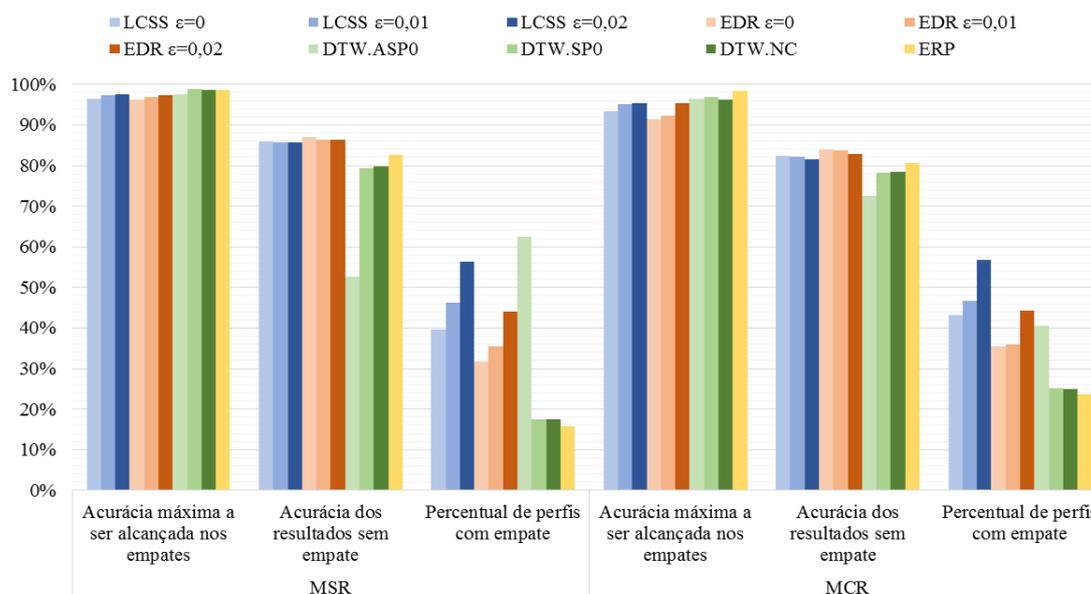
## 6.2. CLASSIFICAÇÃO

A análise inicial foi realizada com a avaliação dos resultados globais das variações das medidas de (dis)similaridade (Figura 25). Na figura cada medida foi representada por uma cor; os LCSS e EDR tem variações de claro para escuro à medida que o parâmetro  $\epsilon$  aumenta; no caso do DTW a variação vai do claro, quando não há restrição, médio para restrição simétrica e escuro para restrição assimétrica. Dos resultados produzidos, notou-

se que além da classificação do perfil para um equipamento específico, também podem ocorrer empates de resultados, ou seja, uma série temporal associada a mais de um equipamento. Dessa forma, foram ser calculadas a acurácia dos resultados com e sem empate, sabendo que *com* o empate o valor do indicador de qualidade era um potencial que *poderia* ser atingido via posterior(s) processo(s) de classificação.

A acurácia, para ambas as situações, apresentou maior parte dos valores acima de 80%, tanto para o modelo MSR quanto para o MCR. O próximo ponto de interesse é o percentual de perfis que precisariam ser submetidos a reclassificações. No caso do modelo MSR, esses perfis foram aqueles que apresentaram resultados com empates. Para o MCR, além desses perfis houve também aqueles perfis que, por conta das restrições sugeridas, não foram enquadrados em nenhum equipamento e, portanto, não passaram para a fase de classificação. Naturalmente, o ideal é que haja a menor quantidade possível de reclassificações. Desse modo, de acordo com o observado na Figura 25 e limitando os empates a percentuais inferiores a 40%, a medida LCSS (com todas as variações de  $\epsilon$ ), o DTW com restrição ASP0 e o EDR  $\epsilon = 0,02l/s$  foram considerados inviáveis tanto no MSR quanto no MCR.

O LCSS ao buscar uma subsequência longa em comum, sem levar em conta as lacunas, tem maior possibilidade de confundir os equipamentos (alto percentual de resultados múltiplos) já que entre equipamentos distintos haviam comportamentos em comum. O DTW com ASP0 teve restrição de movimentos, diferentes dos outros avaliados, e, ao que tudo indicou, influenciou negativamente os resultados. Ao aumentar a tolerância entre vazões ( $\epsilon$ ), tornaram-se mais permissíveis certas associações e ampliou-se muito a possibilidade de distâncias iguais a zero e, portanto, de empates. Entre o MSR e MCR, o primeiro mostrou menor tendência em produzir resultados com empate, logo as análises a seguir, com foco nos equipamentos por residência, foram realizadas com o modelo MSR e com as medidas DTW (NC e SP0), EDR ( $\epsilon = 0l/s$  e  $\epsilon = 0,01l/s$ ) e ERP.



**Figura 25** – Resultados globais de classificação, apresentando a acurácia para resultados com e sem empate, assim como o percentual de perfis com empate de resultados.

## 6.2.1. ANÁLISE POR EQUIPAMENTO

### A. BACIA SANITÁRIA

Em termos de sensibilidade (Tabela 6), houve grande variedade de percentuais, entre 30% e 95%. Avaliando através da moda e mediana, as medidas EDR ( $\epsilon = 0$  l/s e  $\epsilon = 0,01$  l/s) e ERP apresentaram os resultados superiores. Em grande parte dos casos, o percentual em volume foi superior em relação a quantidade de eventos, denotando que foram detectados mais eventos de longa duração. Em geral para a bacia sanitária houve equívocos com torneira interna e chuveiro, possivelmente para aquelas séries temporais mais curtas.

Avaliando a precisão (Tabela 7) em termos de quantidade de eventos, notou-se que o classificador identificou corretamente a maioria dos casos de bacia sanitária. Em aproximadamente metade dos casos que a precisão foi alta para os eventos também foi alta em volume, levando a mesma conclusão que a derivada da sensibilidade, provavelmente os eventos com menor duração foram confundidos com torneiras internas e chuveiros.

Os resultados em geral foram bastante interessantes, porém para a residência B a sensibilidade e precisão foram baixos. Nessa casa, o classificador tende a confundir bacia sanitária com chuveiro. Atribuiu-se isso a semelhança dos comportamentos de ambos os equipamentos; a distribuição de vazões (Figura 8) de ambos era muito parecida, inclusive em duração (Figura 11) e volume (Figura 12). Embora, na hierarquia de equipamentos adotada para essa residência (Quadro 6), a bacia sanitária vinha em primeiro lugar, a presença de vazamentos nela e no chuveiro provavelmente contribuiu para maior índice de equívocos. A casa F também obteve resultados inferiores ao esperado, e isso poderia estar associado aos mesmos problemas que houve na residência B: presença de vazamentos que confundiram os resultados, além de comportamentos comuns com o chuveiro.

**Tabela 6** - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a bacia sanitária. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	67,6%	47,4%	65,6%	46,0%	84,7%	62,7%	87,2%	65,2%	85,4%	62,7%
<b>B</b>	34,5%	25,1%	30,3%	21,1%	31,6%	24,2%	33,4%	25,2%	40,3%	32,7%
<b>C</b>	70,4%	68,3%	67,3%	64,9%	91,7%	88,6%	84,4%	82,5%	87,6%	84,8%
<b>D</b>	67,4%	64,1%	64,1%	61,1%	69,3%	65,3%	72,1%	68,0%	72,2%	68,2%
<b>E</b>	71,1%	50,3%	67,3%	46,5%	85,9%	61,6%	82,0%	60,0%	84,9%	63,0%
<b>F</b>	29,7%	29,6%	27,1%	26,7%	56,4%	55,5%	48,7%	50,0%	47,0%	48,0%
<b>G</b>	67,9%	66,2%	68,1%	65,8%	64,8%	66,9%	65,6%	68,7%	45,7%	53,5%
<b>H</b>	60,1%	57,3%	60,0%	57,5%	94,4%	93,8%	73,7%	70,4%	95,0%	92,8%
<b>I</b>	80,9%	80,7%	78,4%	77,8%	93,7%	94,5%	94,0%	94,5%	93,2%	94,0%
<b>J</b>	37,3%	36,8%	36,0%	35,7%	47,7%	42,3%	61,6%	56,8%	59,4%	56,0%
<b>Média</b>	58,7%	52,6%	56,4%	50,3%	72,0%	65,5%	70,3%	64,1%	71,1%	65,6%
<b>Mediana</b>	67,5%	53,8%	64,9%	52,0%	77,0%	64,0%	72,9%	66,6%	78,5%	62,9%
<b>Desvio</b>	18,0%	18,1%	18,2%	18,3%	21,7%	22,5%	18,6%	18,6%	21,2%	19,9%

**Tabela 7** - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a bacia sanitária. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	73,0%	66,2%	72,2%	63,1%	81,1%	80,4%	82,2%	82,7%	71,1%	70,9%
<b>B</b>	22,6%	17,4%	21,7%	15,6%	24,4%	23,0%	24,3%	26,9%	24,2%	21,0%
<b>C</b>	82,4%	77,6%	82,1%	77,2%	90,4%	91,4%	93,0%	93,5%	83,7%	83,4%
<b>D</b>	54,6%	58,3%	53,0%	55,4%	50,4%	63,0%	54,8%	64,2%	52,0%	57,4%
<b>E</b>	82,0%	86,0%	82,1%	86,9%	84,2%	86,8%	88,8%	90,5%	85,1%	85,9%
<b>F</b>	67,4%	83,0%	67,0%	81,6%	73,9%	82,3%	86,7%	90,1%	86,8%	88,1%
<b>G</b>	78,2%	71,4%	78,9%	71,4%	78,0%	80,8%	77,3%	79,0%	78,4%	75,2%
<b>H</b>	95,8%	91,2%	96,1%	91,6%	90,9%	94,2%	90,1%	93,6%	85,5%	88,4%

<b>I</b>	91,2%	93,0%	94,2%	95,6%	52,8%	76,2%	52,6%	73,8%	80,0%	82,1%
<b>J</b>	65,9%	71,4%	66,6%	74,1%	48,8%	69,5%	52,6%	67,7%	51,9%	58,9%
<b>Média</b>	71,3%	71,5%	71,4%	71,3%	67,5%	74,7%	70,2%	76,2%	69,9%	71,1%
<b>Mediana</b>	75,6%	74,5%	75,6%	75,7%	75,9%	80,6%	79,8%	80,9%	79,2%	78,6%
<b>Desvio</b>	21,0%	22,0%	21,8%	23,1%	22,1%	20,5%	22,8%	20,3%	20,7%	20,9%

## B. CHUVEIRO

O indicador sensibilidade (Tabela 8) mostrou que o valor da média e da mediana foram pouco menores para o chuveiro. Da mesma forma que para a bacia sanitária, a detecção em volume foi superior ao número de eventos, demonstrando que em geral os perfis com volumes e durações maiores foram corretamente classificados. Para as residências A, C, D, E e J o classificador conseguiu identificar boa parte dos usos (Tabela 9). As precisões para as F, G, H e I foram intermediárias. Para o F, usos com maior volume foram confundidos com bacia sanitária e volumes menores com torneira interna; em G e H, o classificador atribuiu usos curtos de chuveiro com torneira interna; na casa I, eventos com volumes altos foram confundidos com torneira externa. A casa B apresentou resultados de precisão e sensibilidades pouco melhores para esse equipamento.

Da mesma forma que a bacia sanitária, os resultados em geral das medidas EDR ( $\varepsilon = 0l/s$  e  $\varepsilon = 0,01l/s$ ) e ERP foram superiores.

**Tabela 8** - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para o chuveiro. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.

	DTW-NC		DTW-SP0		EDR- $\varepsilon=0$ l/s		EDR- $\varepsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	70,1%	49,7%	65,9%	45,8%	81,1%	60,1%	90,4%	72,3%	85,3%	63,1%
<b>B</b>	29,7%	38,4%	29,7%	38,4%	40,7%	50,6%	41,2%	51,3%	40,7%	48,8%
<b>C</b>	28,9%	20,8%	24,5%	18,6%	69,0%	55,4%	72,5%	55,8%	67,9%	54,2%
<b>D</b>	32,0%	27,4%	26,8%	23,9%	43,0%	36,9%	54,4%	46,5%	65,3%	56,4%
<b>E</b>	46,8%	40,5%	44,2%	37,2%	76,2%	63,1%	84,3%	73,6%	77,4%	67,3%
<b>F</b>	45,1%	49,9%	46,1%	47,8%	37,7%	41,5%	49,0%	50,4%	42,5%	46,7%
<b>G</b>	30,4%	30,1%	30,5%	32,0%	59,7%	58,1%	77,4%	69,3%	86,0%	76,5%
<b>H</b>	39,5%	30,3%	36,4%	28,6%	40,7%	33,7%	56,6%	46,4%	51,7%	39,7%
<b>I</b>	85,6%	79,3%	83,1%	77,6%	42,3%	45,4%	45,7%	51,9%	83,7%	63,2%
<b>J</b>	55,6%	47,6%	52,3%	45,1%	62,3%	54,8%	62,4%	57,3%	76,4%	63,7%
<b>Média</b>	46,4%	41,4%	44,0%	39,5%	55,3%	49,9%	63,4%	57,5%	67,7%	58,0%
<b>Mediana</b>	42,3%	39,5%	40,3%	37,8%	51,3%	52,7%	59,5%	53,8%	72,1%	59,7%
<b>Desvio</b>	19,1%	16,7%	18,9%	16,5%	16,4%	10,1%	16,9%	10,4%	17,3%	10,9%

**Tabela 9** - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para o chuveiro. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	85,6%	54,6%	85,8%	55,0%	94,0%	82,7%	90,4%	74,3%	76,8%	64,5%
<b>B</b>	43,6%	37,6%	44,0%	38,4%	45,7%	42,4%	45,4%	41,3%	44,3%	39,3%
<b>C</b>	67,9%	50,4%	60,2%	42,9%	76,3%	76,5%	62,2%	60,2%	62,4%	59,5%
<b>D</b>	82,4%	66,1%	81,6%	64,2%	80,6%	68,4%	86,6%	77,0%	80,6%	68,4%
<b>E</b>	59,3%	43,0%	60,5%	42,2%	78,9%	73,2%	71,1%	64,6%	63,1%	52,6%
<b>F</b>	56,4%	23,1%	57,1%	22,9%	59,5%	31,1%	62,7%	31,7%	55,5%	23,4%
<b>G</b>	39,0%	11,9%	41,4%	12,3%	68,7%	41,8%	67,4%	33,5%	57,1%	25,4%
<b>H</b>	65,3%	45,8%	64,3%	45,3%	67,3%	62,5%	49,9%	47,3%	73,4%	55,3%
<b>I</b>	69,7%	66,6%	69,4%	66,6%	37,1%	47,2%	39,6%	52,4%	62,4%	64,0%
<b>J</b>	88,3%	69,5%	89,6%	73,0%	92,2%	82,6%	92,8%	84,5%	88,7%	71,8%
<b>Média</b>	65,7%	46,9%	65,4%	46,3%	70,0%	60,8%	66,8%	56,7%	66,4%	52,4%
<b>Mediana</b>	66,6%	48,1%	62,4%	44,1%	72,5%	65,5%	65,1%	56,3%	62,8%	57,4%
<b>Desvio</b>	16,8%	19,0%	16,4%	19,2%	18,5%	18,8%	18,8%	18,5%	13,3%	17,4%

### C. TORNEIRA EXTERNA

Esse equipamento foi o que apresentou os menores valores de sensibilidade, tanto para o volume quanto para o número de eventos (Tabela 10). Na residência C mais de 50% do volume conseguiu ser detectado. O grande problema, que se repete nas residências, foi que os usos curtos foram classificados como torneiras internas e usos mais longos foram identificados como chuveiro. Essa afirmação foi baseada na observação das matrizes de confusão dessas residências (não foram apresentadas por conta do grande volume de informações). Avaliando a ordem dos *templates* por residência, não foi observado nexos entre a ordenação e os resultados atingidos.

Todavia, para as residências A, C e E, a capacidade de detecção do classificador é muito boa, chegando a 100% (Tabela 11). Isso mostrou que usos mais específicos de torneira externa são corretamente detectados pelo classificador. Para a casa I, a precisão atingiu valores medianos e para a casa H o classificador praticamente não foi capaz de detectar corretamente os eventos de equipamento.

**Tabela 10** - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira externa. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	42,8%	20,9%	39,5%	17,5%	42,4%	14,1%	44,4%	20,2%	34,6%	34,4%
<b>C</b>	57,6%	32,4%	55,9%	29,6%	51,1%	15,6%	52,8%	19,4%	62,1%	33,2%

<b>E</b>	39,7%	27,3%	38,7%	28,0%	21,3%	15,4%	33,1%	19,2%	28,0%	22,1%
<b>H</b>	8,6%	14,9%	7,5%	13,5%	8,3%	13,3%	0,9%	1,7%	6,7%	8,7%
<b>I</b>	29,0%	17,4%	29,7%	18,5%	13,1%	17,1%	18,2%	24,2%	16,4%	29,2%
<b>Média</b>	35,6%	22,6%	34,2%	21,5%	27,2%	15,1%	29,9%	16,9%	29,6%	25,5%
<b>Mediana</b>	39,7%	20,9%	38,7%	18,5%	21,3%	15,4%	33,1%	19,4%	28,0%	29,2%
<b>Desvio</b>	18,2%	7,2%	17,7%	7,0%	18,7%	1,5%	20,7%	8,7%	21,1%	10,5%

*Tabela 11 - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira externa. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.*

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	55,0%	54,8%	49,9%	49,3%	57,0%	55,2%	81,9%	67,0%	82,1%	74,3%
<b>C</b>	99,1%	98,4%	99,1%	98,2%	100,0%	100,0%	100,0%	100,0%	99,2%	98,4%
<b>E</b>	83,8%	87,5%	83,4%	87,7%	87,2%	87,0%	96,0%	94,3%	93,9%	91,2%
<b>H</b>	4,6%	4,8%	4,1%	4,4%	6,3%	5,1%	0,6%	0,9%	3,1%	2,5%
<b>I</b>	54,1%	31,1%	52,6%	32,6%	45,0%	32,3%	54,1%	40,9%	41,0%	25,6%
<b>Média</b>	59,3%	55,3%	57,8%	54,5%	59,1%	55,9%	66,5%	60,6%	63,9%	58,4%
<b>Mediana</b>	55,0%	54,8%	52,6%	49,3%	57,0%	55,2%	81,9%	67,0%	82,1%	74,3%
<b>Desvio</b>	36,1%	38,8%	36,5%	38,8%	36,9%	38,9%	41,0%	40,9%	40,9%	42,2%

#### D. TORNEIRA INTERNA

Dentre os equipamentos modelados, a torneira interna foi a que apresentou valores altos para ambos os indicadores, na maioria das residências. A sensibilidade (Tabela 12) tanto para o volume quanto para a quantidade de eventos atingiu valores de quase 100%. A casa F apresentou valores entre 50% e 60%, com perfis de volumes maiores atribuídos ao chuveiro. Em termos de precisão (Tabela 13), diferentemente da bacia sanitária e do chuveiro, o percentual do volume foi menor que o da quantidade de eventos, mostrando que o classificador foi capaz de detectar bem usos com volumes menores. Isso ficou bem claro nas residências A e F.

*Tabela 12 - Resultados da sensibilidade por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira interna. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.*

	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	82,7%	92,8%	82,7%	92,6%	94,4%	98,0%	92,9%	96,8%	83,4%	93,1%
<b>B</b>	63,8%	83,4%	67,1%	83,9%	71,8%	89,6%	69,1%	90,7%	58,4%	84,0%
<b>C</b>	95,3%	98,6%	94,7%	98,4%	98,8%	99,8%	98,6%	99,7%	96,4%	99,3%
<b>D</b>	95,2%	98,9%	95,8%	98,8%	99,1%	99,9%	98,6%	99,8%	94,1%	99,1%
<b>E</b>	95,8%	98,6%	96,0%	98,5%	99,4%	99,8%	99,5%	99,8%	97,2%	99,2%
<b>F</b>	54,8%	51,0%	56,7%	51,2%	60,0%	65,2%	53,7%	57,5%	58,0%	54,0%
<b>G</b>	76,9%	86,4%	79,0%	86,2%	82,2%	94,4%	76,5%	91,0%	71,1%	87,1%

<b>H</b>	80,2%	87,7%	79,8%	87,7%	86,8%	92,9%	83,4%	90,6%	78,9%	87,0%
<b>I</b>	79,6%	92,6%	82,2%	93,4%	73,0%	93,7%	62,8%	84,8%	65,3%	89,0%
<b>J</b>	93,5%	97,0%	94,5%	97,6%	95,0%	98,6%	93,9%	97,7%	88,0%	95,5%
<b>Média</b>	81,8%	88,7%	82,9%	88,8%	86,0%	93,2%	82,9%	90,8%	79,1%	88,7%
<b>Mediana</b>	81,5%	92,7%	82,5%	93,0%	90,6%	96,2%	88,2%	93,9%	81,2%	91,1%
<b>Desvio</b>	14,1%	14,4%	13,2%	14,3%	13,8%	10,4%	16,6%	12,8%	15,2%	13,4%

*Tabela 13 - Resultados da precisão por medida de (dis)similaridade (e variações) por volume (V) e quantidade de eventos (Q) por casa para a torneira interna. O dégradé de azul, verde e vermelho mostra cor mais intensa para valores maiores e menos intensa para valores menores.*

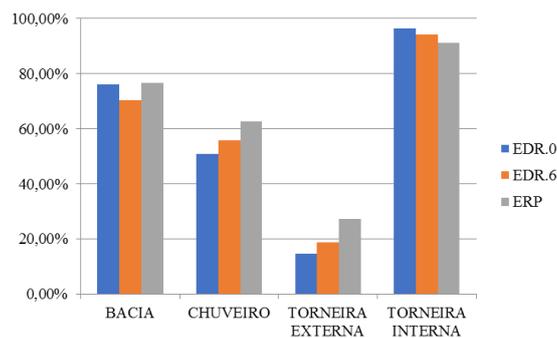
	DTW-NC		DTW-SP0		EDR- $\epsilon=0$ l/s		EDR- $\epsilon=0,01$ l/s		ERP	
	V	Q	V	Q	V	Q	V	Q	V	Q
<b>A</b>	25,3%	55,3%	23,2%	54,1%	35,2%	60,3%	37,2%	63,4%	41,9%	61,8%
<b>B</b>	66,3%	90,0%	63,6%	88,8%	80,0%	94,1%	84,4%	94,4%	89,4%	95,9%
<b>C</b>	61,7%	93,1%	59,1%	92,4%	84,8%	96,9%	83,9%	96,8%	90,2%	97,4%
<b>D</b>	66,0%	95,5%	63,5%	95,2%	82,0%	97,7%	84,0%	97,9%	93,3%	98,8%
<b>E</b>	48,6%	83,2%	44,9%	82,1%	67,1%	88,0%	74,8%	90,2%	70,3%	87,3%
<b>F</b>	25,3%	67,1%	25,6%	65,5%	23,7%	66,4%	21,7%	63,7%	27,6%	69,8%
<b>G</b>	68,7%	93,9%	69,2%	93,9%	74,2%	95,7%	81,2%	96,5%	83,9%	95,9%
<b>H</b>	53,8%	86,6%	52,4%	86,4%	73,2%	92,1%	77,1%	89,7%	77,2%	91,1%
<b>I</b>	53,0%	88,0%	48,4%	87,0%	65,3%	93,1%	69,8%	89,7%	94,0%	94,6%
<b>J</b>	68,0%	89,3%	66,8%	89,0%	79,3%	93,1%	80,1%	93,3%	84,0%	94,0%
<b>Média</b>	53,7%	84,2%	51,7%	83,4%	66,5%	87,7%	69,4%	87,6%	75,2%	88,7%
<b>Mediana</b>	57,8%	88,7%	55,8%	87,9%	73,7%	93,1%	78,6%	91,7%	83,9%	94,3%
<b>Desvio</b>	16,5%	12,9%	16,4%	13,3%	20,6%	13,2%	21,9%	13,0%	22,8%	12,6%

## 6.2.2. ELEIÇÃO DA MEDIDA DE (DIS)SIMILARIDADE IDEAL

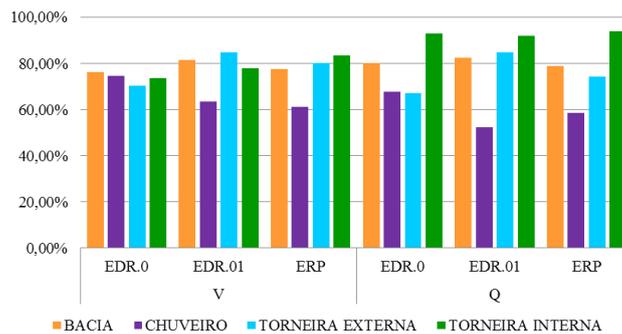
Sintetizando as análises anteriores, pôde ser percebido que na maioria dos casos a medida DTW (NC e SP0) apresentou desempenho inferior as outras medidas em avaliação. As medidas de melhor desempenho apresentaram resultados bastante interessantes, mas que não foram comuns a todas as residências. Pôde-se depreender que as diferenças de resultados ocorreram por dois motivos básicos. O primeiro foi relacionado com a dinâmica residencial de consumo que, com base na seção de análise dos dados, mostrou-se bastante particularizada por residência, por ser influenciada pelos costumes dos moradores, que abarcam rotinas, nível de conscientização do consumo, condições das instalações hidráulicas e dos tipos de equipamentos disponíveis. Por outro lado, relatos do grupo de pesquisa apontaram que o Trace Wizard® poderia realizar classificações equivocadas e, nesse caso, a principal consequência foi a alocação de perfis em classes que não eram as verdadeiras. Esses acontecimentos causaram distanciamento entre a estrutura de comportamento dos equipamentos no banco de dados classificado e

os padrões estabelecidos na semana investigativa. Na sequência, foi aplicada à parte majoritária dos dados ao MSR testando as medidas ERP e EDR ( $\varepsilon = 0l/s$  e  $\varepsilon = 0,01l/s$ ).

Seguindo os critérios de análise mencionados anteriormente, notou-se dificuldade em eleger a medida melhor representativa. Optou-se então por analisar os indicadores através da mediana, que por sua característica representou a maioria das casas. Avaliando os valores da mediana da sensibilidade para o volume (Figura 26a) observou-se que a medida ERP se destacou para o chuveiro e torneira externa, mas para a bacia sanitária os resultados foram empatados com o EDR ( $\varepsilon = 0l/s$ ) e ligeiramente inferior aos EDR ( $\varepsilon = 0l/s$  e  $\varepsilon = 0,01l/s$ ) para a torneira interna. Do ponto de vista da precisão (Figura 26b), também faltou clareza tanto do volume quanto da quantidade de eventos para efetuar a seleção. Por outro lado, se fez necessário lembrar do percentual de perfis com empate. De acordo com a Tabela 14, tanto para a análise inicial quanto para a análise final, o ERP mostrou menor quantidade de eventos com empate e a importância disso advém de que ele conseqüentemente conseguiu captar mais volumes que as outras medidas. Isso foi fator significativo, desde que o objetivo foi captar a maior quantidade de volume consumido. Assim, o ERP foi eleito como medida de dissimilaridade que melhor se adequou.



(a)



(b)

**Figura 26** - Mediana da sensibilidade para o volume (a) e mediana da precisão para volume V e número de eventos Q (b).

**Tabela 14** - Percentual de perfis com empates para análise inicial e para a parte majoritária dos dados.

Medida de similaridade	DADOS	
	Análise inicial	Análise final
<b>EDR.0</b>	31,9%	31,5%
<b>EDR.01</b>	35,5%	34,8%
<b>ERP</b>	15,8%	15,5%

Os resultados finais com uso do ERP foram sumarizados nas Tabelas 15 e 16. Os resultados da amostra para a maior massa de dados não variaram tanto, no sentido de que as casas que estavam com problemas nos em certos índices ainda permaneceram (a amostra de fato representou o todo).

A média foi calculada para comparação com outros trabalhos na intenção de verificar se havia viabilidade do algoritmo proposto. Comparando com base na precisão, que é o indicador que tem correlatos entre todas as referências (vide Quadro 1), na média tanto a bacia sanitária quanto a torneira interna atingiram resultados próximos e até um pouco superiores a algumas ferramentas. O chuveiro teve um desempenho médio um

pouco abaixo dos outros e não haviam torneiras externas na lista de equipamentos desses trabalhos.

*Tabela 15 - Sumário dos resultados de precisão por volume (V) e quantidade de eventos (Q) para todas as residências utilizando o modelo MSR e a medida ERP.*

	BACIA		CHUVEIRO		TORNEIRA EXTERNA		TORNEIRA INTERNA	
	V	Q	V	Q	V	Q	V	Q
<b>A</b>	81,88%	57,55%	86,28%	66,24%	30,53%	35,60%	81,02%	93,31%
<b>B</b>	40,32%	33,13%	34,72%	44,09%			56,96%	83,70%
<b>C</b>	87,04%	84,88%	72,46%	60,92%	56,95%	31,89%	97,77%	99,59%
<b>D</b>	71,52%	67,13%	68,27%	59,52%			93,19%	99,09%
<b>E</b>	84,37%	59,95%	74,36%	66,77%	26,13%	21,03%	96,91%	99,12%
<b>F</b>	45,97%	46,61%	43,85%	46,30%			56,75%	54,30%
<b>G</b>	49,95%	57,20%	82,71%	70,51%			68,77%	85,58%
<b>H</b>	94,69%	92,23%	49,13%	39,28%	4,73%	9,25%	80,61%	87,78%
<b>I</b>	94,68%	94,82%	83,27%	64,14%	14,46%	27,25%	62,29%	88,74%
<b>J</b>	60,36%	56,49%	78,22%	65,93%			87,29%	95,21%
<b>Média</b>	71,1%	65,0%	67,3%	58,4%	26,6%	25,0%	78,2%	88,6%
<b>Mediana</b>	76,7%	58,8%	73,4%	62,5%	26,1%	27,2%	80,8%	91,0%

*Tabela 16 - Sumário dos resultados de precisão por volume (V) e quantidade de eventos (Q) para todas as residências utilizando o modelo MSR e a medida ERP.*

	BACIA		CHUVEIRO		TORNEIRA EXTERNA		TORNEIRA INTERNA	
	V	Q	V	Q	V	Q	V	Q
<b>A</b>	76,90%	74,19%	62,14%	57,57%	80,05%	74,29%	42,75%	62,08%
<b>B</b>	22,92%	21,70%	39,41%	34,60%			88,87%	95,55%
<b>C</b>	85,11%	86,28%	66,47%	65,65%	100,00%	100,00%	89,26%	97,24%
<b>D</b>	54,75%	60,37%	79,75%	67,75%			93,60%	98,84%
<b>E</b>	83,45%	83,67%	56,53%	49,08%	94,23%	92,93%	70,90%	86,68%
<b>F</b>	86,51%	87,92%	54,39%	23,31%			27,77%	69,34%
<b>G</b>	75,41%	72,80%	53,26%	21,28%			83,93%	95,98%
<b>H</b>	79,39%	85,73%	78,22%	59,87%	2,76%	2,95%	75,88%	90,63%
<b>I</b>	78,13%	84,93%	60,25%	63,38%	38,46%	24,90%	94,57%	94,15%
<b>J</b>	58,43%	63,97%	86,94%	68,61%			82,84%	93,70%
<b>Média</b>	70,1%	72,2%	63,7%	51,1%	63,1%	59,0%	75,0%	88,4%
<b>Mediana</b>	77,5%	78,9%	61,2%	58,7%	80,1%	74,3%	83,4%	93,9%

### 6.3. TÉCNICAS AUXILIARES PARA AMPLIAÇÃO DO DESEMPENHO DO CLASSIFICADOR

Um aspecto interessante a ser levado em consideração é a agregação de técnicas auxiliares no intuito de melhorar os índices de qualidade do algoritmo de classificação. O primeiro alvo de melhoria possivelmente seria a quantidade de perfis referência. O caráter

abrangente de um conjunto de perfis padrão perpassa pelo conhecimento das dinâmicas nas residências, que promove a capacidade de afirmar até que ponto os perfis obtidos são suficientes. Wonders et al. (2016) propõe a síntese de dados para melhorar a qualidade do classificador e Vašak et al. (2015) também propõe o aumento no banco de dados de treinamento, mudando a proporção treinamento/teste. Em ambos os casos houve aumentos nos índices. Para este projeto, estas duas propostas são equivalentes a aumentar o banco de dados de referência. Desde que há diferença temporal entre o esse trabalho e o período de obtenção dos dados, não seria possível retornar as residências e encontrar as mesmas condições. Logo, o caminho seria simular mais perfis ou trazer dados do banco para as referências. No primeiro caso, a síntese teria um caráter mais complexo desde que a variabilidade de tipos de uso é grande. Ao transportar dados do banco de treinamento, constatou que além do já sabido aumento do tempo de processamento, a matriz de confusão apresentou mais erros.

De outra forma, Nguyen et al. (2013a) propõem o uso de características de consumo para encaminhar os perfis teste para os equipamentos mais prováveis, além de empregar informações sobre probabilidade de uso a depender do horário. O uso de informações de tendência de consumo por equipamento tem o inconveniente de precisar ser individualizado, porque nem todas as residências têm os mesmos equipamentos, a vazão de água fornecida pode variar entre localidades (diversificando o perfil de consumo), os equipamentos automatizados variam de consumo entre marcas e programas e o comportamento dos moradores pode não ser similar. No caso deste trabalho, numa forma análoga foi aplicada restrição de volume e duração através do modelo MCR, mas com pouco sucesso. Também se tentou empregar a probabilidade de uso de certo equipamento dado o horário e o dia, mas para um mesmo horário, a probabilidade entre equipamentos é muito parecida e em aplicando, os indicadores não evidenciaram crescimento. É necessário então estudar outras técnicas que demonstrem aderência para o problema em questão e que auxiliem no processo de classificação.

#### 6.4. ALGORITMO INTEGRADO

Após os resultados da seção de separação de usos simultâneos e classificação de séries temporais de consumo, a integração dessas foi realizada estudando-se duas hipóteses. Os resultados gerais constam na Tabela 17, nela o percentual de perfis de

consumo em comum com a classificação do Trace Wizard® atingiu o valor de mais de 80% em ambos os casos, com destaque para a hipótese 2. A acurácia da classificação ficou acima de 75%, mais uma vez com destaque para a hipótese 2. Numa visão mais detalhada, observando as matrizes de confusão (Tabelas 18 e 19), a sensibilidade e a precisão não variaram muito de um modelo para outro. Concluindo, considerando todos os pontos, a hipótese 2 foi a mais indicada.

*Tabela 17 – Resultados acurácia dos modelos sugeridos.*

	HÍPOTESE 1	HÍPOTESE 2
<b>Percentual de perfis em comum com Trace Wizard®</b>	82%	85%
<b>Acurácia da classificação</b>	76%	78%

*Tabela 18 - Matriz de confusão para a hipótese 1.*

		PREDITO					
		Bacia	Chuveiro	Tanque	Torneira	NA	<b>SENSIBILIDADE</b>
REAL	Bacia	72	12	0	2	5	79,12%
	Chuveiro	6	12	0	2	1	57,14%
	Tanque	7	0	4	5	0	25,00%
	Torneira	1	0	0	897	1	99,78%
PRECISÃO		83,72%	50,00%	100,00%	99,01%	-	

		PREDITO					
		Bacia	Chuveiro	Tanque	Torneira	NA	<b>SENSIBILIDADE</b>
REAL	Bacia	79	13	0	2	5	79,80%
	Chuveiro	6	16	0	3	1	61,54%
	Tanque	8	0	7	7	0	31,82%
	Torneira	1	1	0	920	1	99,67%
PRECISÃO		84,04%	53,33%	100,00%	98,71%	-	

Embora os algoritmos tenham sido aplicados e analisados por residência, numa proposta mais ampla, essa abordagem se torna trabalhosa. O ideal seria estudar mais a fundo as características socioeconômicas que possam estar mais associadas ao consumo d'água e propor agrupamentos entre casas que compartilhem padrões de consumo semelhantes, facilitando a logística do processo.

## 7. CONCLUSÃO

O intento desse projeto foi a construção de uma ferramenta de reconhecimento de usos de equipamentos hidráulicos residenciais, cujo diferencial foi a consideração dos diferentes comportamentos de uso por equipamento e a capacidade de realizar separações de eventos simultâneos sem intervenção do pesquisador.

Em termos de algoritmo de classificação, o uso de medidas de (dis)similaridade aplicadas num modelo análogo em conceito ao *k-nearest neighbor* demonstrou a capacidade de contemplar as variações de uso para um mesmo de equipamento e concomitantemente atingir bons resultados. A dificuldade de acerto em certos momentos deveu-se a comportamentos comuns entre equipamentos distintos, evidenciados no uso da medida LCSS que produziu grande percentual de empate. Em geral, o aumento do parâmetro  $\epsilon$ , tanto para o LCSS quanto para o EDR, aumentou a flexibilidade de semelhança entre vazões, mas isso intensificou a produção de resultados com empate. Dentre as variações da medida DTW, a ASP0 apresentou desempenho inferior, tanto na qualidade de classificação quanto no alto percentual de empates. De acordo com os critérios de sensibilidade, focando no volume, e de precisão obtiveram-se os EDR ( $\epsilon = 0l/s$  e  $\epsilon = 0,01l/s$ ) e ERP como possíveis para utilização, mas com base na produção de resultados com empate, o ERP destacou-se e foi eleito como o mais indicado. Em termos de sensibilidade com relação ao volume, a bacia sanitária, chuveiro e torneira interna obtiveram respectivamente médias de 71,1%, 67,3% e 78,2%, sendo que para mais da metade das residências conseguiu-se alcançar mais de 70% de detecção entre esses equipamentos. Para torneira externa esse indicador obteve valores abaixo do esperado com média de 26,6%, por conta principalmente de ser confundido com a torneira interna. Em termos de precisão com relação ao número de eventos, o classificador conseguiu para a bacia sanitária, chuveiro, torneira externa e torneira interna médias de respectivamente 72,2%, 51,1%, 59,0%, 88,4%; em algumas casas a capacidade de detecção para a torneira externa chegou a mais de 90%. Com base na precisão relacionada ao número de eventos (indicador com correlatos entre as todas as referências (Quadro 1)), em média tanto a bacia sanitária quanto a torneira interna atingiram resultados próximos e até um pouco superiores a algumas ferramentas. O chuveiro teve um desempenho médio um pouco abaixo dos outros e não houve correlato à torneira externa na lista de equipamentos desses trabalhos.

Quanto ao algoritmo de separação, idealizaram-se as etapas de análise da simultaneidade (para rotular eventos paralelos), corte vertical (para eventos contíguos) e corte horizontal (para eventos sobrepostos). Constatou-se que a premissa para o corte vertical não apresentou o mesmo nível de abrangência que o corte horizontal, logo foi desconsiderada. Após a avaliação da variação de parâmetros, os valores finais para a tolerância (*tol*) e percentual de aceitação da moda (*pma*) foram respectivamente, 0,011/s e 90%. Com esses, o percentual de acerto foi superior a 80% na maioria das residências, isso sem a intervenção do pesquisador. Cabe mencionar que o algoritmo construído teve um caráter adaptável a sistemas de medição com resoluções e intervalos de medição distintos, além de ser maleável ao poder serem escolhidas as etapas de corte a serem aplicadas e os valores dos parâmetros.

Após analisar as hipóteses de integração dos algoritmos, concluiu-se que algoritmo de separação seguido da classificação foi a configuração ideal. Cabe ratificar que o resultado atingido no algoritmo integrado não sofreu interferência humana, mostrando o potencial que pôde ser atingido.

A ferramenta pode ser melhorada via aplicação de técnicas auxiliares ao algoritmo de classificação. Nos trabalhos avaliados, houveram aplicações de informações e técnicas para auxiliar na melhoria dos indicadores, fosse numa pré-seleção para melhor encaminhamento, fosse com análises posteriores para usos não identificados. Houve a tentativa nesse trabalho de aumentar o banco de dados de referência através da transferência de dados do banco classificado pelo Trace Wizard®, no entanto além do aumento de tempo computacional causado, os indicadores não mostraram crescimento. A aplicação da matriz de probabilidade por hora por dia também não auxiliou no processo de pré-seleção.

A ferramenta construída teve sua motivação na cooperação ao mecanismo de gestão da demanda, que pode torna-se mais efetiva com informações de natureza mais detalhada. Espera-se que todo o desenvolvimento realizado auxilie na difusão de métodos para trabalhar dados de medidores inteligentes, a partir dos resultados obtidos nesse trabalho e nas publicações advindas dele. Com o avanço nos estudos nessa direção, haverá aperfeiçoamento na qualidade da resposta e a criação de interface amigável facilitará a interpretação dos resultados.

## 8. SUGESTÃO DE TRABALHOS FUTUROS

- A análise de vazamento seria importante adição ao processo. Essa necessita ser mais dinâmica com o fim de informar ao morador, em tempo real, de problemas no sistema hidráulico;
- A construção de interface amigável é outro passo necessário para tornar a ferramenta proposta aplicável;
- Como todo algoritmo que pode ser melhorado, sugestões de técnicas auxiliares são oportunos para a classificação, assim como a tentativa de desenvolver um algoritmo autômato, capaz de aprender com o tempo e, portanto, se especializar;
- Em se tratando de uma ferramenta que demanda uma mudança na forma como a medição do consumo é realizada, seria interessante desenvolver uma análise de viabilidade da implantação do medidor inteligente, explanando pontos como limitações, logística e prioridades.

## 9. PUBLICAÇÕES

AGUIAR FILHO, A. M.; MELLO, M. S.; OLIVEIRA-ESQUERRE, K. P.; BOTELHO, G. L. P.; KIPERSTOK, A.. Identificação de perfil de consumo de água domiciliar por equipamento hidráulico e modo de uso via fuzzy-clustering como ferramenta de gestão de recursos hídricos no ambiente urbano. In: CONGRESSO ABES/FENASAN 2017, 29., 2017, São Paulo. *Anais...* São Paulo: ABES, 2017.

MELLO, M.; OLIVEIRA-ESQUERRE, K.; BOTELHO, G.; KIPERSTOK A.. Comparative study of similarity measures used to classify residential water flow pattern of low-income households in Salvador - Brazil. **Computer Aided Chemical Engineering**, v. 44, p. 1405- 1410, 2018.

## REFERÊNCIAS

ANA. **Conjuntura dos recursos hídricos no Brasil**: Informe 2016. Brasília: Agência Nacional de Águas, 2016. 95 p.

BEAL, C.; STEWART, R. A. **South East Queensland residential end use study**. Goal Coast: Urban Water Security Research Alliance, 2011. 174 p. (47).

BENNETT, C.; STEWART, R. A.; BEAL, C. D. ANN-based residential water end-use demand forecasting model. **Expert systems with applications**, v. 40, n. 4, p. 1014-1023, 2013.

BÔAS, A. L. V. **Uso racional de água e eficiência energética em habitações de interesse social**. Racionalização do consumo de água na habitação de interesse social: relatório de atividades. Salvador, 2014. 06 p., Coordenação de Asher Kiperstok.

BOTELHO, G. L. P. **Avaliação do consumo de água em domicílios**: Fatores intervenientes e metodologia para setorização dos usos. 2013. 197 f. Dissertação (Mestrado) - Curso de Engenharia Ambiental, Mestrado em Meio Ambiente, Águas e Saneamento - MAASA, Universidade Federal da Bahia, Salvador, 2013.

BOTELHO, G. L. P. **Gestão da demanda por água em residências de baixa e média renda – BA - “projeto GESAGUA”**: relatório de atividades. Salvador, 2014. 35 p., Coordenação de Karla Patricia Santos Oliveira Rodríguez Esquerre.

BOTELHO, G. L. P. **Uso racional de água e eficiência energética em habitações de interesse social**. Racionalização do consumo de água na habitação de interesse social: relatório de atividades. Salvador, 2014. 35 p., Coordenação de Asher Kiperstok.

CHEN, L.; NG, R. On the marriage of lp-norms and edit distance. In: PROCEEDINGS OF THE THIRTIETH INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 30., 2004, Toronto. *Anais...* VLDB Endowment, 2004, p. 792-803.

CHEN, L.; ÖZSU, M. T.; ORIA, V. Robust and fast similarity search for moving object trajectories. In: PROCEEDINGS OF THE 2005 ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2005, Nova Iorque. *Anais...* ACM, 2005. p. 491-502.

COMINOLA A. et al. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. **Environmental Modelling & Software**, v. 72, p. 198-214, 2015.

CRUZ, L. N. da. **Uso racional de água e eficiência energética em habitações de interesse social**. Racionalização do consumo de água na habitação de interesse social: relatório de atividades. Salvador, 2014. 13 p., Coordenação de Asher Kiperstok.

FONTDECABA, S. et al. An approach to disaggregating total household water consumption into major end-uses. **Water resources management**, v. 27, n. 7, p. 2155-2177, 2013.

JESUS, L. S. de. **Uso racional de água e eficiência energética em habitações de interesse social**. Racionalização do consumo de água na habitação de interesse social: relatório de atividades. Salvador, 2014. 13 p., Coordenação de Asher Kiperstok.

JORGENSEN, B.; GRAYMORE, M.; O'TOOLE, K. Household water use behavior: An integrated model. **Journal of environmental management**, v. 91, n. 1, p. 227-236, 2009.

KEOGH, E.; RATANAMAHATANA, C. A. Exact indexing of dynamic time warping. **Knowledge and information systems**, v. 7, n. 3, p. 358-386, 2005.

KURBALIJA, V. et al. The influence of global constraints on similarity measures for time-series databases. **Knowledge-Based Systems**, v. 56, p. 49-67, 2014.

LIU, A. et al. Detailed water-use feedback: A review and proposed framework for program implementation. **Utilities Policy**, v. 43, p. 140-150, 2016.

MORRISON, J.; FRIEDLER, E. A critical review of methods used to obtain flow patterns and volumes of individual domestic water using appliances. **Urban Water Journal**, v. 12, n. 4, p. 328-343, 2015.

NGUYEN, K. A.; STEWART, R. A.; ZHANG, H. An autonomous and intelligent expert system for residential water end-use classification. **Expert Systems with Applications**, v. 41, n. 2, p. 342-356, 2014.

NGUYEN, K. A.; ZHANG, H.; STEWART, R.A. Application of Dynamic Time Warping algorithm in prototype selection for the disaggregation of domestic water flow data into end use events. In: Proceedings of the 34th World Congress of the International Association for Hydro-Environment Research and Engineering: 33rd Hydrology and Water Resources Symposium and 10th Conference on Hydraulics in Water Engineering. Engineers Australia, 2011. p. 2137.

NGUYEN, K.; STEWART, R. A.; ZHANG, H. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. **Environmental modelling & software**, v. 47, p. 108-127, 2013b.

NGUYEN, K.; ZHANG, H.; STEWART, R. A. Development of an intelligent model to categorise residential water end use events. **Journal of Hydro-environment Research**, v. 7, n. 3, p. 182-201, 2013a.

SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. **IEEE transactions on acoustics, speech, and signal processing**, v. 26, n. 1, p. 43-49, 1978.

SANTOS, A. L. de O. A. **Uso racional de água e eficiência energética em habitações de interesse social**. Racionalização do consumo de água na habitação de interesse social: relatório de atividades. Salvador, 2014. 15 p., Coordenação de Asher Kiperstok.

STUDART, T. M. C. e CAMPOS, N. Gestão da Demanda. In: CAMPOS, N.; STUDART, T. M. C. **Gestão de Águas: Princípios e Práticas**. 2. ed. Fortaleza: ABRH - Associação Brasileira de Recursos Hídricos, 2001. Cap. 5. p. 63-80.

VAŠAK, M.; BANJAC, G.; NOVAK, H. Water Use Disaggregation Based on Classification of Feature Vectors Extracted from Smart Meter Data. **Procedia Engineering**, v. 119, p. 1381-1390, 2015.

VLACHOS, M.; KOLLIOS, G.; GUNOPULOS, D. Discovering similar multidimensional trajectories. In: PROCEEDINGS 18TH INTERNATIONAL

CONFERENCE ON DATA ENGINEERING, 18., 2002, San Jose. *Anais...* IEEE, 2002. p. 673-684.

WANG, X. et al. Experimental comparison of representation methods and distance measures for time series data. **Data Mining and Knowledge Discovery**, v. 26, n. 2, p. 279-309, 2013.

WONDERS, M.; GHASSEMLOOY, Z.; HOSSAIN, M. A. Training with synthesised data for disaggregated event classification at the water meter. **Expert Systems with Applications**, v. 43, p. 15-22, 2016.

## APÊNDICE A

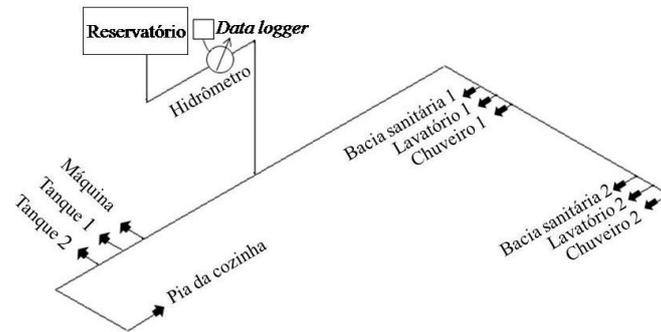


Figura 27 - Esquema de instalação do novo sistema de medição em uma das casas em estudo (BOTELHO, 2013).

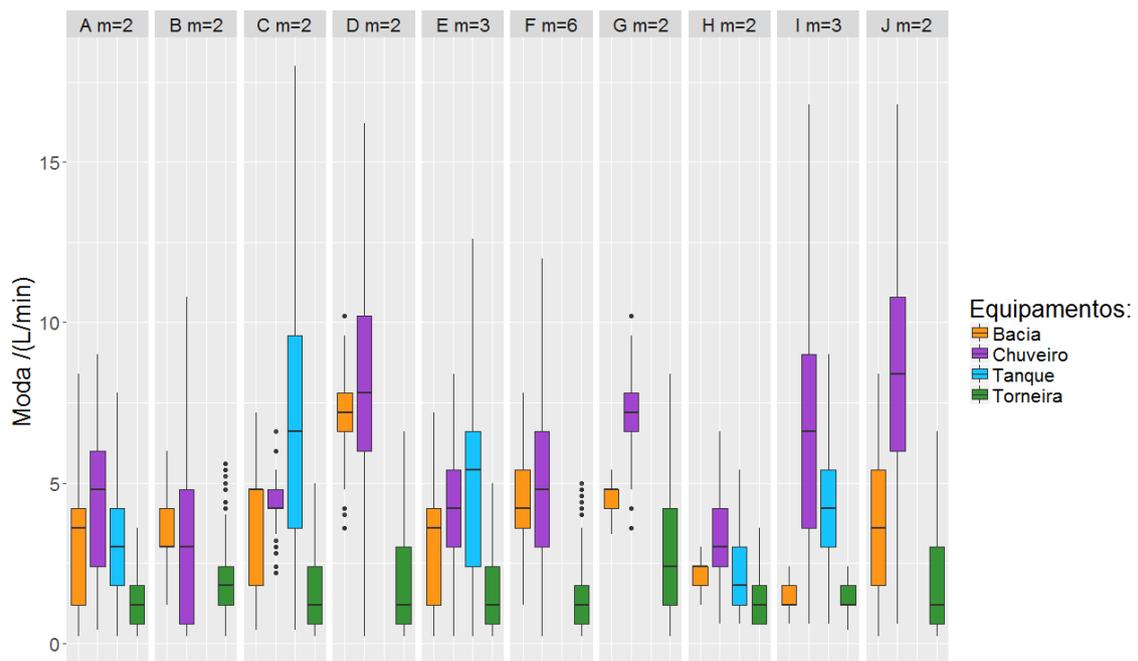


Figura 28 – Gráfico boxplot da moda (L/min) para os equipamentos por residência.

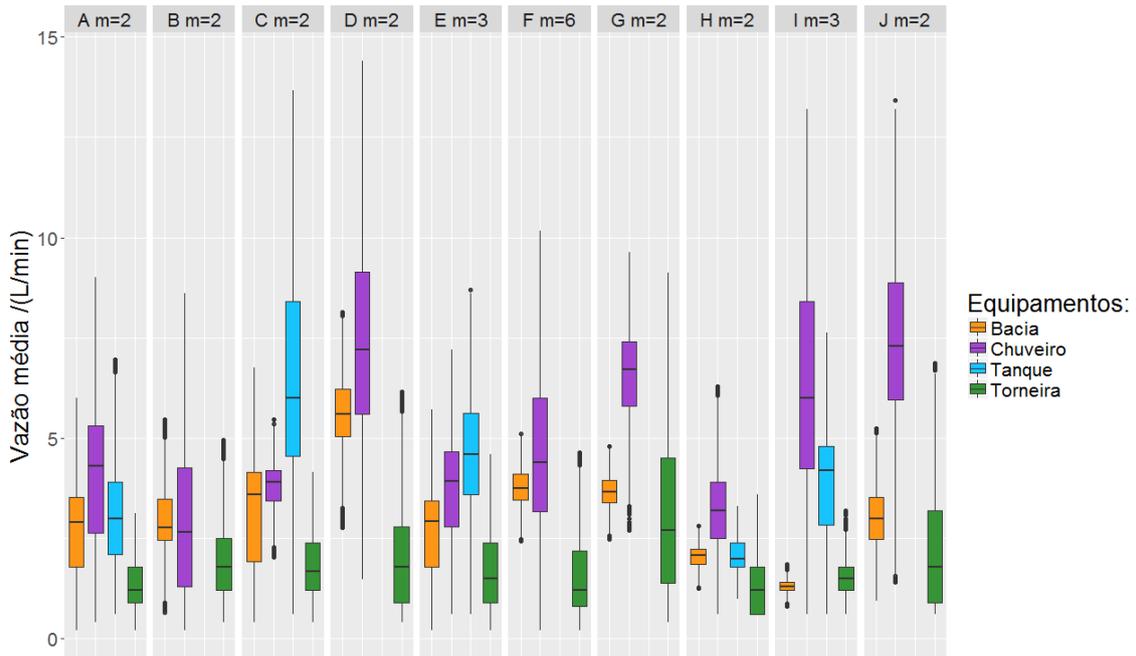


Figura 29 - Gráfico boxplot da vazão média (L/min) para os equipamentos por residência.

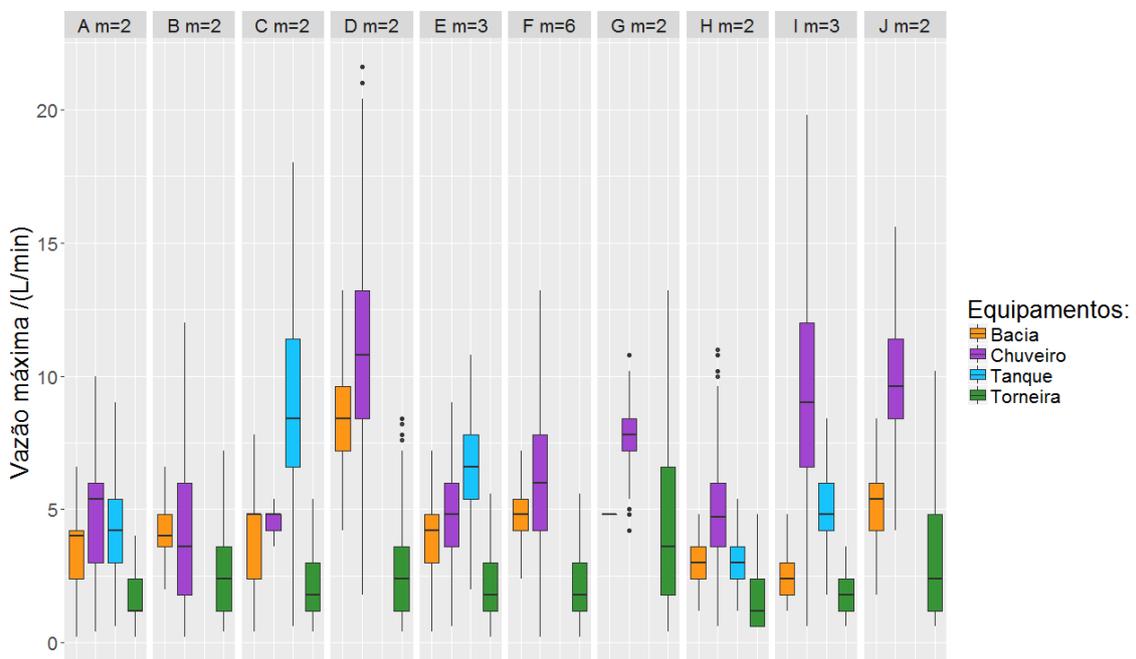


Figura 30 - Gráfico boxplot da vazão máxima (L/min) para os equipamentos por residência.

## APÊNDICE B

*Quadro 6 - Relação dos templates aplicados por residência. Ao lado de cada template consta o percentual de vezes que foi aplicado.*

CASA	ORDEM	TEMPLATE 1	70%	TEMPLATE 2	30%
A	1	Bacia		Vazamento	
	2	Chuveiro		Bacia	
	3	Torneira externa		Chuveiro	
	4	Torneira interna		Torneira externa	
	5	-		Torneira interna	
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>90%</b>	<b>TEMPLATE 2</b>	<b>7%</b>
B	1	Vazamento		Vazamento	
	2	Bacia		Bacia	
	3	Chuveiro		Chuveiro	
	4	Chuveiro/Vazamento		Torneira interna	
	5	Torneira interna		-	
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>76%</b>	<b>TEMPLATE 2</b>	<b>24%</b>
C	1	Bacia		Bacia	
	2	Chuveiro		Chuveiro	
	3	Torneira externa 1		Filtro	
	4	Filtro		Torneira externa	
	5	Torneira interna 1		Torneira interna	
	6	Torneira externa 2		-	
	7	Torneira interna 2		-	
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>100%</b>		
D	1	Vazamento			
	2	Bacia			
	3	Chuveiro			
	4	Torneira interna			
	5	Torneira externa			
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>69%</b>	<b>TEMPLATE 2</b>	<b>31%</b>
E	1	Máquina de lavar		Máquina de lavar	
	2	Bacia		Bacia	
	3	Chuveiro		Chuveiro	
	4	Torneira externa		Torneira externa	
	5	Torneira interna		Torneira interna	
	6	Vazamento		-	
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>87%</b>	<b>TEMPLATE 2</b>	<b>13%</b>
F	1	Vazamento		Vazamento	
	2	Bacia		Bacia	
	3	Chuveiro		Chuveiro	
	4	Torneira interna		Torneira interna	
	5	-		Torneira externa	
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>100%</b>		
G	1	Vazamento			
	2	Máquina de lavar			
	3	Chuveiro			
	4	Bacia			
	5	Torneira interna			
	6	Torneira externa			
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>100%</b>		
H	1	Vazamento			
	2	Chuveiro			
	3	Bacia			
	4	Torneira externa			
	5	Torneira interna			
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>100%</b>		
I	1	Vazamento			
	2	Torneira interna			
	3	Torneira externa			
	4	Bacia			
	5	Chuveiro			
	<b>ORDEM</b>	<b>TEMPLATE 1</b>	<b>96%</b>	<b>TEMPLATE 2</b>	<b>3%</b>
J	1	Bacia		Vazamento	
	2	Chuveiro		Máquina de lavar	
	3	Torneira interna		Chuveiro	
	4	Torneira externa		Bacia	
	5	Vazamento		Torneira interna	
	6	-		Torneira externa	

**UFBA**  
**UNIVERSIDADE FEDERAL DA BAHIA**  
**ESCOLA POLITÉCNICA**

**PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA INDUSTRIAL - PEI**

Rua Aristides Novis, 02, 6º andar, Federação, Salvador BA

CEP: 40.210-630

Telefone: (71) 3283-9800

E-mail: [pei@ufba.br](mailto:pei@ufba.br)

Home page: <http://www.pei.ufba.br>

