

**IV-279 - 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**

Adelmo Menezes de Aguiar Filho⁽¹⁾

Graduando em Engenharia Química pela Universidade Federal da Bahia (UFBA). Possui formação técnica em operação de processos industriais e químicos pelo Instituto Federal da Bahia (IFBA).

Mariza Souza de Mello⁽¹⁾

Atualmente é graduada em Engenharia Química da Universidade Federal da Bahia. Participa do Grupo de Pesquisas GAMA - UFBA, focada na área de análise estatística multivariada aplicada. Mestranda no Programa de Pós-Graduação em Engenharia Industrial na UFBA.

Karla Patrícia Santos Oliveira Rodríguez⁽¹⁾

Engenheira Química pela Universidade Federal de Alagoas (UFAL). Mestre e Doutora em Engenharia Química pela Universidade Estadual de Campinas (UNICAMP). Pós-doutora em Engenharia Sócio-Ambiental pela Hokkaido University (Hokudai, Japão). Professora do Departamento de Engenharia Química e do Programa de Pós Graduação em Engenharia Industrial da Escola Politécnica, Universidade Federal da Bahia.

Gabriella Laura Peixoto Botelho⁽²⁾

Engenheira Sanitarista e Ambiental, Mestre em Meio Ambiente, Águas e Saneamento e Doutoranda em Engenharia Industrial pela Universidade Federal da Bahia. Professora Assistente na Universidade Federal do Recôncavo da Bahia.

Asher Kiperstok⁽¹⁾

Engenheiro Civil pela TECHNION, Israel Institute of Technology. MSc. e PhD em Engenharia Química, Tecnologias Ambientais pela University of Manchester Institute of Science and Technology (UMIST, Inglaterra). Coordenador da Rede de Tecnologias Limpas e Minimização de Resíduos (TECLIM) e do Programa de Pós Graduação em Produção Limpa – Departamento de Engenharia Ambiental, (EP/UFBA).

Endereço⁽¹⁾: Rua Aristides Novis, 02 - Universidade Federal da Bahia - Departamento de Engenharia Química – Federação - Salvador - BA - CEP: 40210-630 - Brasil - Tel.: (71) 9-9209-5169 - e-mail: adelmo.aguiar.filho@gmail.com

RESUMO

A oportunidade de conhecimento mais apurado sobre o consumo d'água residencial possibilita grande avanço no desenvolvimento de políticas de melhor distribuição e preservação d'água, além de consequente redução da produção de efluentes. Atualmente, informações quanto ao consumo de água por residência conta com dados mensais provenientes do hidrômetro, que fornece volumes mensais de uso, contudo o mesmo não permite verificação da frequência e modo de uso dos equipamentos hidráulicos a partir dos consumidores. Essa limitação pode ser suprimida com o uso de medidores inteligentes pois propiciam uma visão mais precisa através da medição em intervalos curtos de tempo (minutos, segundos), produzindo perfis aproximados do consumo. Assim, com base nessas informações, este trabalho teve o objetivo de produzir um modelo de identificação de equipamentos a partir dos dados de perfis de uso não agregados; avaliar a regularidade de uso e volume de consumo por equipamento e a partir desses propor indicadores de uso d'água mais adequados à realidade local. Os dados foram provenientes de residências do bairro Chapada do Rio Vermelho, Salvador, por representarem a faixa de baixa renda, que é predominante na cidade. O modelo de reconhecimento de padrão foi construído com base na técnica de análise de componentes principais (PCA) e do algoritmo *fuzzy c-means* probabilístico. Os resultados indicaram considerável dificuldade na classificação dos perfis por conta da incapacidade do algoritmo de criar grupos que separassem os equipamentos sem confundir-los. A limitação apresentada se deu pela natureza dos dados pois, uma vez que os equipamentos variam entre semiautomáticos e manuais, há uma multiplicidade de usos que torna os perfis variados e, por conseguinte, com grande espalhamento, dificultando a criação de subespaços vetoriais com mínimo possibilidade para associações equivocadas. No entanto, apesar dos resultados obtidos, este estudo tem sua importância na exposição dessa característica dos dados, que direciona os próximos estudos a modelos que abarquem essas peculiaridades e possibilitem uma avaliação mais precisa do comportamento local.

PALAVRAS-CHAVE: Gestão de recursos hídricos, Desagregação do fluxo de água residencial, Clusterização *fuzzy c-means*.

INTRODUÇÃO

O caráter não renovável da água potável e sua aplicabilidade diversa e em diferentes seguimentos instigam a necessidade de melhor entendimento acerca do seu uso para desenvolver e aperfeiçoar estratégias de gestão da demanda. No quesito estudo da demanda residencial, há diversos trabalhos que buscam relacionar o volume d'água utilizado com características dos moradores e o contexto local no qual estão inseridos. Cominola et al. (2015) esperam que haja grande crescimento urbano mundial, que por consequência levará ao aumento demanda residencial, a qual hoje cobre uma grande parte do abastecimento público de água potável. Já Jorgensen et al. (2009) asseguram que problemas como aumento da população, poluição da água, desenvolvimento urbano, irrigação agrícola, mudanças climáticas e secas contribuem para disparidades entre a disponibilidade de fontes de água de qualidade e demanda de consumo. Complementando, Vašak et al. (2015) afirmam que as autoridades estão enfrentando problemas de escassez de água em muitas partes do mundo e que a redução desse consumo deve ser feita com a informação de como e quando a água é empregada. Assim, aprofundar o estudo sobre o consumo de água potável nesse setor possibilitará uma melhor gestão desse recurso, além de indiretamente contribuir para a redução de gastos energéticos necessários ao processo de tratamento do esgoto, e diminuição do impacto ambiental ocasionado pelo descarte dos efluentes de forma inadequada.

Os estudos relacionados à caracterização do uso de água intradomiciliar foram favorecidos pelas informações 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 consumo de água, energia ou gás natural em intervalos de tempos curtos, com emissor de sinal integrado, unidos a equipamentos de captação automática e quase instantânea do sinal emitido. Os dados captados fornecem um perfil de consumo mais preciso e, portanto, torna-se uma importante fonte de informações acerca do consumo interno. Liu et al. (2016) mostram que, para melhorar a conservação da água utilizada em residências urbanas, é necessário que o consumidor conheça seu consumo interno de forma detalhada e o uso de medidor inteligente, tecnologia recente, tem uma contribuição muito importante nisso. Nguyen et al. (2014) e Bennett et al. (2013), com preocupações similares, têm seus estudos focados no consumo residencial de água na Austrália, que há certo tempo vem sofrendo com secas, e ressaltam que o acesso a dados fornecidos por medidores inteligentes foi um facilitador para modelagem do consumo por aparelhos. Neste trabalho, os perfis de consumo foram utilizados num modelo de identificação de padrões.

No contexto deste trabalho, a importância de identificar qual ou quais aparelhos (uso simultâneo) estão sendo utilizados decorre da possibilidade de conhecer, de modo mais detalhado, como os moradores direcionam o uso d'água e, por conseguinte, quais os aparelhos mais utilizados e seus volumes associados, além do volume diário residencial consumido. Embora campanhas de estímulo a conservação da água já sejam utilizadas, o feedback à sociedade sobre o uso desagregado d'água é mais preciso e pode promover conservação a nível individual e residencial (LIU et al, 2016); espera-se que o conhecimento gerado neste trabalho possa futuramente ser levado ao consumidor final e assim permitir uma gestão individual e consciente do consumo da água.

Através da obtenção de resultados de classificação consistentes é possível derivar informações de grande utilidade, e dessas, novos indicadores de gestão do recurso, para ambas concessionárias e consumidores. Liu et al. (2016) propõem um aplicativo de interface amigável, o HWU (*Home Water Update*), no qual se podem 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, comparação do 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, fornecer dicas personalizadas de como reduzir o consumo. Vašak et al. (2015), com o COP (*Customer On-line Portal*), apresentam 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) têm 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 deve incluir médias de consumo

por localidades, para comparações de acordo com característica(s) relevante(s); padrões demanda diurna diária; relatórios de previsão de demanda de água para regiões selecionadas com base em dados fornecidos em tempo real. Eles ainda ressaltam que tais informações podem auxiliar no desenvolvimento de programas de conservação mais direcionados e também na modelagem de redes de tubulações otimizadas.

O estudo, com foco em residências de baixa renda, foi realizado no bairro Chapada do Rio Vermelho, Salvador e os dados foram obtidos através da instalação de medidor inteligente e *data logger* por unidade. A escolha deste bairro deveu-se ao fato do mesmo representar o perfil da população de baixa renda, que é uma parcela expressiva da população de Salvador, no aspecto socioeconômico. Sendo assim, conhecer o consumo de água residencial dessa parte da população é de grande importância. O sistema de monitoramento captou o consumo intradomiciliar a cada dez segundos, entre dezembro de 2010 a abril de 2014. Em comparação com os aparelhos hidráulicos residenciais utilizados por classes com rendas superiores, nota-se que aqui existe um predomínio de aparelhos de uso manual em detrimento dos automáticos, que são mais caros e despendem custos extras de energia elétrica. Essa característica torna esse estudo mais desafiador por conta dos diversos usos possíveis dos equipamentos.

O objetivo deste trabalho é desenvolver uma ferramenta de identificação do uso e modo de uso de equipamentos hidráulicos intradomiciliar com base no algoritmo *fuzzy c-means* probabilístico como forma de aprimoramento da gestão do consumo de água em centros urbanos, possibilitando a proposição de indicadores adequados à realidade socioeconômica local, o acompanhamento de seus hábitos de consumo e o desenvolvimento de políticas públicas de maior efetividade na redução de perdas e desperdício.

MATERIAIS E MÉTODOS

Coleta de dados

Medições da vazão requerida a cada 10 segundos foram registradas para um total de 10 residências localizadas no bairro da Chapada do Rio Vermelho na cidade do Salvador durante o período de dezembro de 2010 até abril de 2014 (Tabela 1). A medição foi realizada via instalação de registradores eletrônicos configurados para armazenar pulsos de resolução igual a 0,1 L/pulso, sendo acoplados a um *data logger* para armazenamento dos dados. A instalação do equipamento de medição é realizada na tubulação diretamente conectada a caixa d'água da residência.

Tabela 1. Organização do plano de amostragem.

ID da residência	Número de residentes	Número de pontos de água	Início da coleta	Fim da coleta	Dias monitorados
H1	2	12	07/11/2011	31/01/2012	66
H2	2	4	13/12/2012	19/12/2013	328
H3	2	7	22/12/2010	24/01/2012	316
H4	2	5	13/12/2012	24/04/2014	412
H5	3	12	17/12/2010	18/10/2011	189
H6	6	9	09/04/2013	18/12/2013	244
H7	2	8	28/07/2013	31/03/2014	230
H8	2	5	14/04/2013	31/03/2014	316
H9	4	5	13/12/2012	24/04/2014	355
H10	2	5	14/07/2013	31/03/2014	257

Os dados colhidos foram então desagregados em seus usos finais através de consultas aos moradores ou do uso do software TraceWizard® para posterior treinamento do modelo proposto conforme procedimento detalhado em Botelho (2013). Destaca-se o caráter semiautomático do processo, uma vez que o uso do software se fez acompanhado de posterior avaliação da razoabilidade de suas saídas, que quando não atendida tinha sua resposta corrigida pela avaliação de um especialista ou contato direto com o morador da residência.

Para fins de predição o uso da bacia sanitária, chuveiro, pia da cozinha e a pia do banheiro, foram os usos finais considerados para o modelo, uma vez que correspondem aos equipamentos hidráulicos com maior frequência em todas as residências simultaneamente. A Tabela 2 apresenta o total de observações categorizadas por equipamento hidráulico de cada residência.

Tabela 2. Quantidade amostrada de usos finais por residência.

ID da residência	Bacia Sanitária	Chuveiro	Pia do banheiro	Pia da cozinha
H1	1456	744	382	2855
H2	3538	4607	158	33
H3	4574	1646	17365	17988
H4	2005	2748	63	397
H5	2843	1622	7113	9060
H6	3518	5245	21	84
H7	1741	1493	0	0
H8	1311	2420	34	251
H9	2216	3801	8827	0
H10	1665	1537	92	791

Foi objeto de interesse, ainda durante a etapa de coleta de dados, a análise de consistência dos dados obtidos através da remoção de *outliers*, detecção de vazamentos nas residências e o cruzamento dos dados desagregados com a medição realizada pela companhia de abastecimento de água responsável (EMBASA).

Modelo de clusterização por Fuzzy C-Means probabilístico

O *Fuzzy C-means (FCM)* consiste em um algoritmo de aprendizado não supervisionado para a identificação de padrões. Neste, uma métrica de similaridade é aplicada às observações de um conjunto de dados para alocação destes em um número pré-definido de grupos. Por sua característica probabilística cada observação é associada com diferentes pertinências aos grupos formados, sendo alocado àquela com maior grau de associação (DUDA O.; HART E.; STORK G., 2000). Computacionalmente, a finalidade do algoritmo é minimizar a função objetivo apresentada na Equação 1.

$$\min J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2 \quad (m \geq 1) \quad \text{equação(1)}$$

na qual u_{ik} corresponde à pertinência do perfil de consumo k ao grupo i ; x_k é o vetor de características do perfil de consumo k e v_i corresponde a localização do grupo; m é denominado fuzzificador, variável associada ao grau de incerteza do agrupamento.

A aplicação desse método consiste inicialmente na determinação do número de grupos (tipos de equipamentos hidráulicos) e na estimativa inicial dos centros desses grupos. O algoritmo minimiza de forma iterativa a função objetivo (Equação 1) resultando no cálculo da pertinência de cada perfil de consumo a um tipo de equipamento hidráulico. Os cálculos necessários foram realizados usando o software RStudio e a linguagem aberta de programação R com auxílio do pacote de funções `e1071` (MEYER, 2017).

A avaliação da qualidade do modelo de clusterização foi realizada via verificação do erro de classificação, através da curva característica de operação do receptor (ROC) e da matriz confusão. A leitura da ROC é feita tomando como critério a razão de falsos positivos (RFP) desejável e verificando o valor correspondente para a razão de verdadeiros positivos (RVP). Desta forma, um agrupamento é dito ideal quando seu RVP é igual à unidade para um RFP nulo (ZOLGHADRI; MANSOORI, 2007). Na matriz de confusão são apresentados o número de observações de um determinado grupo que foram previstas e seu respectivo grupo real.

Determinação das características relevantes para identificação dos perfis de consumo

Dado que os perfis de consumo coletados por equipamento apresentam diferentes durações, a implementação do algoritmo *fuzzy c-means* necessita do pré-tratamento do banco de dados para a obtenção de características representativas (*features*) das séries temporais dos equipamentos, de forma a representar cada perfil com um mesmo número de variáveis. A seleção do algoritmo FCM se destaca por sua flexibilidade, precisão e robustez a ruídos (AGHABOZORGI et al., 2015).

A seleção das *features* foi realizada através da análise de componentes principais (PCA), técnica de redução de dimensionalidade linear que preserva a estrutura de correlação entre as variáveis do processo (CHIANG et al., 2001). Por meio desta, diferentes *features* foram calculadas a partir dos dados e, sua relevância para

classificação dos perfis de consumo, avaliada a partir da análise dos *loadings* do PCA. Os *scores* obtidos por meio deste procedimento foram utilizados como as entradas do modelo *fuzzy*.

RESULTADOS E DISCUSSÃO

A análise de componentes principais (PCA) quando aplicada ao conjunto de dados, demonstrou através de seus *loadings* (Tabela 1), as *features* mais relevantes na caracterização dos equipamentos hidráulicos. Como as duas primeiras componentes principais mostram-se suficiente para explicar mais de 90% da variabilidade dos dados, verifica-se que os *loadings* mais expressivos são o volume total, o número de repetições da moda e a vazão máxima, resultado concordante com Vařak et al. (2015).

Através desta redução de dimensionalidade, os grupos formados a técnica de clusterização via FCM podem ser interpretados graficamente em duas dimensões. O conhecimento da importância das variáveis se traduz na posterior compreensão de falsos-positivos e aprimoramentos do modelo selecionado.

Tabela 1. *Loadings* obtidos pela técnica de componentes principais.

Feature	Componente principal					
	1°	2°	3°	4°	5°	6°
Repetições da moda	-0.523	0.644	-0.559			
Volume total	-0.727		0.687			
Moda da vazão	-0.212	-0.378	-0.235	0.552	0.669	
Média da vazão	-0.188	-0.328	-0.188		-0.450	-0.783
Vazão Máxima	-0.275	-0.447	-0.281	-0.767	0.205	0.122
Mediana da vazão	-0.207	-0.368	-0.216	0.316	-0.554	0.607

A Figura 1 apresenta as observações do banco de dados escalonadas para a primeira e segunda componentes principais juntamente com os agrupamentos gerados pela técnica *fuzzy c-means* (representada pela distribuição normal bivariada cujos parâmetros de média e variância foram obtidos pelas observações classificadas como pertencentes daquele agrupamento). Tanto os agrupamentos quanto as observações mostraram-se de difícil separação linear, tornando elevado o número de falsos-positivos e enviesando a decisão do modelo para a categoria com maior presença no banco de dados.

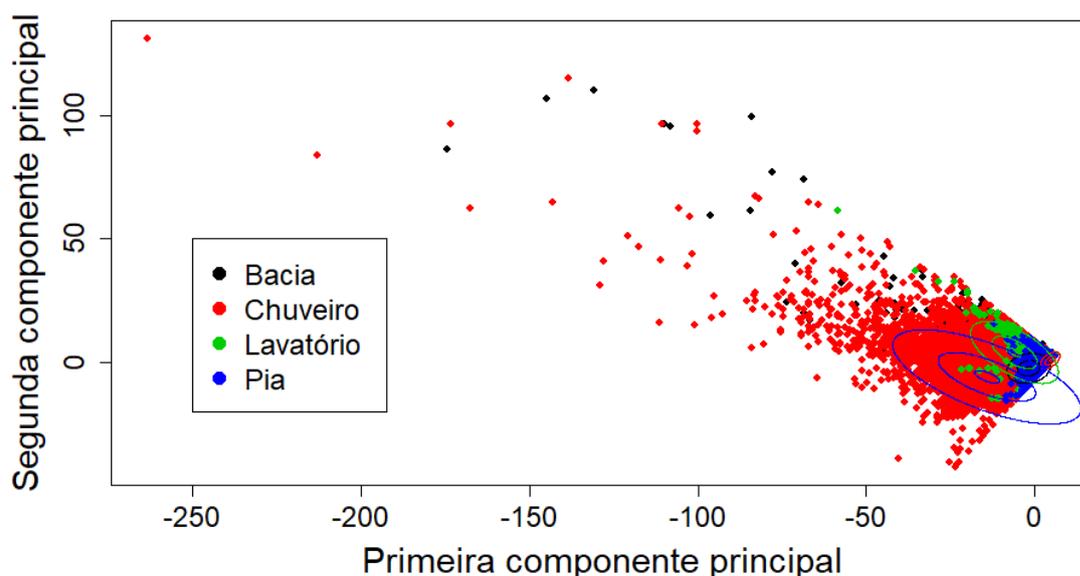


Figura 1. Análise gráfica dos agrupamentos formados pelo *fuzzy-c-means*.

A configuração apresentada na Figura 1 torna a resolução do problema de clusterização via técnicas lineares (*support vector machines*, análise discriminante linear etc.) de abordagem deficiente, uma vez que não existe subespaço vetorial capaz de gerar uma separação satisfatória das categorias de equipamento hidráulica consideradas. Este fato pode ser observado na Tabela 2, que apresenta a matriz de confusão da classificação realizada pelo modelo FCM, no qual verifica-se que o modelo classifica com um grau de acerto variando de 0,05% (pia) até 70% (bacias). Exatidão inferior ao de modelos encontrados na literatura com mesmo propósito (NGUYEN, 2014; VAŠAK, 2015).

As razões para esta configuração podem estar associadas ao modo de uso dos equipamentos hidráulicos por cada morador ou mesmo ao consumo intrínseco ao equipamento utilizado. Como forma de avaliar a influência destes dois aspectos na qualidade do modelo, o conjunto de dados foi segregado em função do morador e técnica de *fuzzy c-means* aplicada individualmente. A Figura 2 apresenta os agrupamentos gerados por residência.

Tabela 2. Matriz de confusão da clusterização realizada.

		Predito			
		Bacia	Chuveiro	Lavatório	Pia
Observado	Bacia	10072	3797	248	58
	Chuveiro	11453	11049	955	4297
	Lavatório	2336	3361	33092	27401
	Pia	1081	7696	11	4

A segregação realizada por residência levou a resultados com equivalente exatidão, culminando em uma classificação ineficaz do ponto de vista da modelagem proposta. Portanto, modelos como o FCM, por sua natureza linear, mostram-se pouco apropriados. Este resultado mostra-se importante para a compreensão da natureza dos dados com os quais está sendo lidado, visto que na literatura modelos de natureza linear foram propostos com exatidão satisfatório como foram os casos dos trabalhos em Vašak (2015), DeOreo (2000), entre outros discutidos em Morrison (2015).

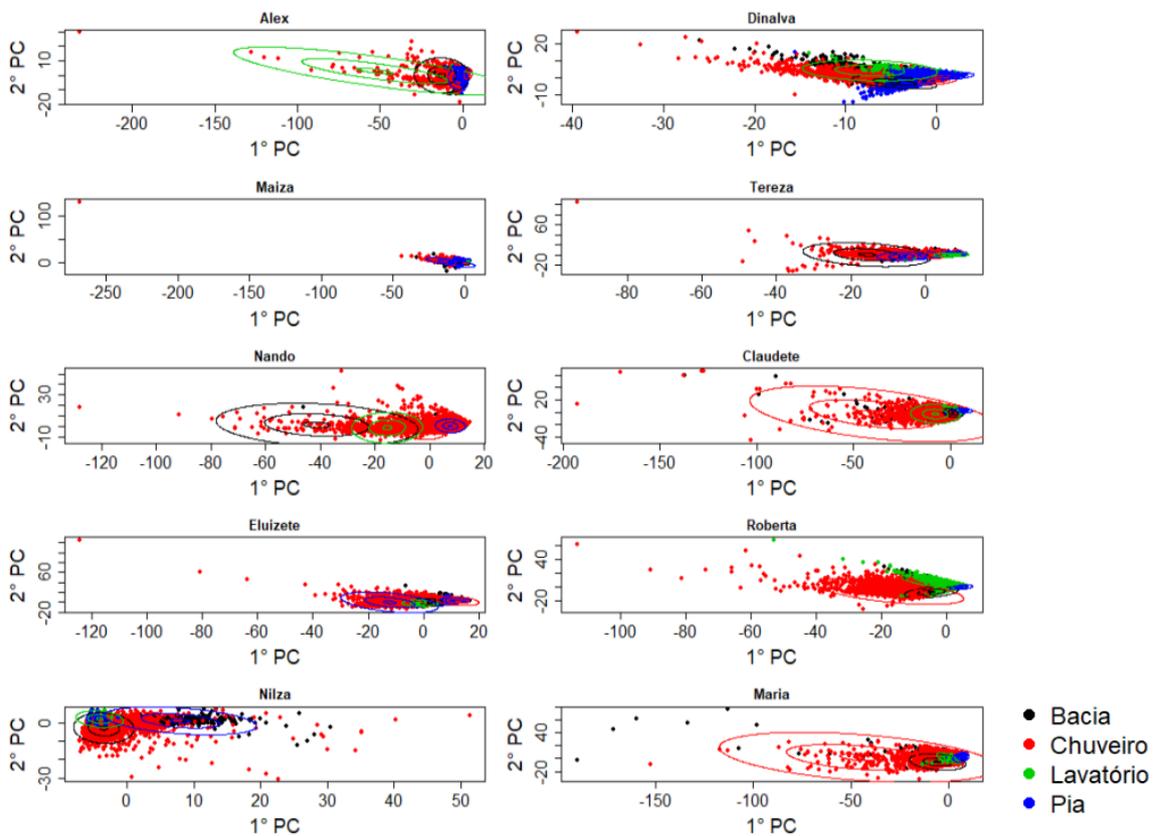


Figura 2. Análise gráfica dos agrupamentos formados pelo *fuzzy c-means*.

Os países em que os trabalhos citados anteriormente foram realizados possuem como característica padrões bem estabelecidos quanto ao consumo de água pelos equipamentos, por exemplo lava-louças, pias e bacias sanitárias com acionador, etc. Isto garante padrões de caracterização dos equipamentos mais claros, evitando, como foi apresentado, o confundimento do uso de uma torneira em banheiros com o uso de um chuveiro.

Com relação à proposta de recomendação de indicadores mais apropriados, no contexto dessa pesquisa, além dos possíveis indicadores citados anteriormente (BENNETT et. al., 2013; NGUYEN et. al., 2014; LIU et. al., 2016), acredita-se poder explorar outras como a avaliação do uso do chuveiro (classificar entre uso prolongado e curto), visando otimizar o tempo de banho; o mesmo para as torneiras, também no sentido de conter o uso excessivo e o fornecimento de informação a respeito de bacias mais economizadoras, com base nos modelos observados.

CONCLUSÕES

Diante de cenários de racionamento e escassez, o conhecimento do comportamento dos indivíduos quanto ao uso da água por equipamento pode suprir informações-chave para a construção de políticas públicas, projetos de renovação de equipamentos e identificação de vazamentos intradomiciliares. No presente trabalho, buscou-se a aplicação da técnica *fuzzy c-means* em um extenso conjunto de dados com o objetivo de identificar os equipamentos a partir dos perfis de consumo em habitações localizadas na cidade do Salvador, Bahia. Apesar de não satisfatório, por conta das características de lineares da técnica FCM, foi possível compreender acerca da natureza dos dados, norteando à exploração de outras técnicas que sejam capazes de lidar com as dificuldades encontradas.

AGRADECIMENTOS

Os dados utilizados nesta pesquisa foram coletados durante os projetos Gestão da Demanda por Água em Residências de Baixa e Média Renda – BA (GESAGUA-BA) - CTHIDRO / Edital MCT/CNPq/CT-Hidro nº 21/2009 (556114/2009-4) e Tecnologias Poupadoras de Água Previsão da Demanda de Água para a Região Metropolitana de Salvador - MCT/CNPq - N ° 14/2012 - Universal (484895/2012-5). Ao CNPq e a todos os pesquisadores que participaram deste projeto, nossos agradecimentos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. 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.
2. BOTELHO, G. L. P. (2013). Avaliação do consumo de água em domicílios: Fatores intervenientes e metodologia para setorização dos usos, 227.
3. COMINOLA, A.; GIULIANI, M.; PIGA, D.; CASTELLETTI, A.; RIZZOLI, A. E. 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.
4. MEYER. D. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. <https://CRAN.R-project.org/package=e1071>.
5. DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. John Wiley & Sons, 2012.
6. 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.
7. LIU, A.; GIURCO, D.; MUKHEIBIR, P. Urban water conservation through customised water and end-use information. *Journal of Cleaner Production*, v. 112, p. 3164-3175, 2016.
8. 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.
9. 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.