Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 820 – 827

Procedia Engineering

www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

Assessment of water use efficiency in the household using cluster analysis

Catarina Jorge[a]*, Paula Vieira[a], Margarida Rebelo[a], Dídia Covas[b]

[a]*National Laboratory for Civil Engineering, Lisbon, Portugal*
[b]*Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*

Abstract

Concerns over the efficiency of household water use have been increasing worldwide: water is intensively consumed with high inefficiencies. Detailed methodologies for the assessment of consumers' behaviour that include the level of efficiency comparatively to their peers (i.e., similar socio-demographic characteristics) are needed to increase public awareness and to develop strategies related to the efficient use of water. This paper presents a methodology for evaluating the efficiency of indoor domestic water use based on peer comparison. Peer groups were established through cluster analyses according to their relevant socio-demographic characteristics. Finally, the most efficient consumers were identified and water efficiency levels were determined for individual households.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Water use efficiency; household; domestic consumption; cluster analysis; socio-demographic variables

Nomenclature

H_E	Household efficiency level [%]
Cl	Cluster average or minimum [litres]
C_{CD}	Consumption based on consumers' data [litres]

* Corresponding author. Tel.: +351 912 549 675.
E-mail address: cnjorge@lnec.pt

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.945

1. Introduction

The influence of socio-demographic characteristics in the domestic water consumption has been analysed and observed by many authors [1,2,3]. Some of these studies report that the householders are enabled to compare their water consumption profile with other neighbouring households in the same district, street, building or DMA. However, this type of comparisons does not lead to accurate results, since consumers characteristics can be very distinct [4]. According to [5], any strategy of water demand management needs the collaboration of the population involved, as it is important to know characteristics such as residence area, family dimension, presence or absence of children and/or elders in the household, income level and instruction, among others, that may influence the use of water.

Another important question to be answered is "if water conservation is more likely when individuals believe that water is scarce or when they perceive that other consumers are also conserving water" [5].

In several studies, surveys have been conducted to obtain socio-demographic information of each household, which allows clustering consumption by varying demographic indicators [6,7]. These relations are very similar to what is found for the efficiency of energy use, since socio-demographic variables are also considered to influence energy consumption.

Regarding the few studies that do exist on this matter [8-10], it seems that older people tend to spend less water *per capita* than the younger. Moreover, families with children and teenagers are expected to use more water, but mainly in external uses [8]. However, [9] research also shown that older people tend to spend more time at home which leads to higher water consumption. Within the demographic variables, the number of people living in the household appears to be the most important driver of consumption [10].

The current research work presented herein is part of a comprehensive methodology to assess the overall households' water use efficiency, which also comprises an evaluation based on efficient patterns and on the performance evaluation of water use devices [11]. These three analyses should be applied in a complementary way to assess most aspects of water efficiency in the household. Results obtained with this research contribute to promote a more efficient use of water and, concurrently, to support water utilities in the optimisation of water supply systems, thereby increasing the services' sustainability and the provided quality.

2. Methodology

The developed peer comparison methodology aims to propose a novel type of efficiency assessment, in which consumers are compared with each other through their own consumption data. This methodology is a four-step procedure as presented in Fig.1.

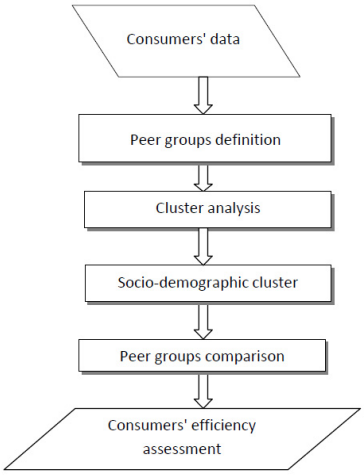

Fig.1. Peer comparison methodology.

The comparison is performed within well-defined peer groups. *Peer groups* are established through a cluster analysis based on their consumption (weekly and *per capita* consumption) and relevant socio-demographic characteristics (i.e., property type, family dimension and family composition by age). In cluster analysis, each individual is selected as a target and is compared to all the other individuals. Based on this comparison, individuals with similar characteristics are chosen as a *peer group*. Then, the event (or behaviour) of the peer group is summarized by each subsequent point in time, and the event (or behaviour) of the target is compared to the summary of the peer group [12].

Correlation matrices are built to support the decision about the variables that most correlated with each other. If the correlation coefficient is high, the variables are not independent and one of them should be discarded from the cluster analysis. This high value shows that variables have large multicollinearity and should not be simultaneously considered as independent variables. In this situation, cluster analysis is not carried out with these variables, because conclusions would not be reliable, since a variable directly influences the other.

The statistical data analysis can be performed with the STATISTICA® and SPSS® software.

Finally, the identification of the most efficient consumers is made through the comparison of their own consumption with the minimum and the average values within their cluster and by calculating an efficiency level for individual households.

3. Case Study

This methodology was applied to 43 households, mostly located in the Lisbon urban area and surrounding municipalities, corresponding to approximately 100 participants. All households have extensive measurements and detailed event records of all indoor water uses, and were surveyed to collect information on socio-demographic variables and on the existing water devices in the households (WC, shower, taps, dishwasher and washing machine).

Table 1 summarizes the overall characteristics of the case study. Property type in this case study varies between T1 (i.e., a single bedroom property) and T4 (i.e, four bedroom properties). Inhabitants' number varies between 1 and 5. Inhabitants' age ranges from less than 1 year and 75 years, with an average age of 35 years. The *per capita* consumption varies between 50 and 286 litres/(person.day), with an average of 138 litres/(person.day). The weekly consumption varies between 473 and 9966 litres/(household.week), with an average of 3014 litres/(household.week).

Table 1. Overall case study characteristics.

	Minimum	Average	Maximum
Socio-demographic characteristic			
Property type	T1	(-)	T4
Family dimension	1	3	5
Family composition by age (years)	0.6	35	75
Consumption characteristic			
Weekly consumption litres/(household.week)	473	3014	9966
Per capita consumption litres/(person.day)	50	138	286

4. Results

4.1. Correlation matrix for socio-demographic variables

As a first step to create the peer groups for consumers' comparison, a parametric correlation analysis was carried out in order to determine which variables are more correlated. The developed correlation analysis considered the following socio-demographic variables:

- family dimension;
- family composition by age;
- property type.

Table 2 shows the correlation matrix obtained. The strongest correlations were observed for the variables: property type and family dimension. Family dimension and family composition by age have also shown a negative correlation but with a low linear association.

Table 2. Correlation matrix for the socio-demographic variables.

Correlation Matrix				
		Family dimension	Family composition by age	Property type
Family dimension	Pearson Correlation	–	-.361*	.470**
	p-value	–	.018	.001
	N	–	43	43
Family composition by age	Pearson Correlation	–	–	.072
	p-value	–	–	.647
	N	–	–	43
Property type	Pearson Correlation	–	–	–
	p-value	–	–	–
	N	–	–	–

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

The variables with a strong correlation – property type and family dimension – were removed from cluster analysis, since one directly influences the other. Therefore, the socio-demographic clusters were built with one or two variables not strongly correlated.

The following clusters were initially considered:

- family dimension clusters;
- property type clusters;
- family composition by age and family dimension clusters;
- family composition by age and property type clusters.

The last cluster was not considered in subsequent analysis, since the differences among cases in the different clusters were not significant and reliable (F_{41}=0.2; p=0.65).

4.2. Family dimension clusters

Four significantly different clusters were obtained based on family dimension characteristics.

Table 3 shows the characteristics of these clusters. Clusters 1 and 2 correspond to small families. The first cluster contains five households with one member and the second cluster includes eight households with two members. Cluster 3 corresponds to larger families with three and four members and includes twenty seven households. Finally, Cluster 4 is formed by the three households with the largest family dimension in this sample (i.e., more than five members).

The major difference between the family dimension clusters is due to weekly consumption, since *per capita* consumption does not reliably vary among the four clusters. This is an expected result because usually when the number of persons in the household increases, the household weekly consumption increases as well. Cluster 1 has an average weekly consumption of 994 litres/(household.week) and Cluster 4 has an average consumption of 5684 litres/(household.week).

Table 3. Characteristics of the family dimension clusters.

Cluster	Number of households	Family dimension (n. of family members)	Per capita consumption [litres/(person.day)]			Weekly consumption [litres/(household.week)]		
			Minimum	Average	Maximum	Minimum	Average	Maximum
Cluster 1	5	1	68	142	286	473	994	2000
Cluster 2	8	2	72	147	196	1004	2047	2747
Cluster 3	27	3, 4	50	135	259	1400	3378	9966
Cluster 4	3	5	127	139	158	4294	5684	8000

4.3. Property type clusters

Table 4 shows the characteristics of the property type clusters. Cluster 1 corresponds to small properties and is composed by nineteen households of the typology T1 and of typology T2 households. The Cluster 2 is composed of larger properties, including twelve T3 households. The last cluster is composed by the largest properties in the sample and includes twelve T4 households.

Similarly to family dimension clusters, the major difference between the three property type clusters is the weekly consumption. The correlation matrix (Table 2) shows a high linear association between property type and family dimension, which justifies the results obtained to both variables, since households with a higher number of residents correspond, in general, to larger dwellings.

Table 4. Characteristics of the property type clusters.

Cluster	Number of households	Property type	Per capita consumption [litres/(person.day)]			Weekly consumption [litres/(household.week)]		
			Minimum	Average	Maximum	Minimum	Average	Maximum
Cluster 1	19	T1, T2	50	134	286	473	2206	4311
Cluster 2	12	T3	74	148	259	2062	3363	5437
Cluster 3	12	T4	55	135	236	894	3943	9966

4.4. Family composition by age and family dimension clusters

Table 5 presents the characteristics obtained for the two clusters of family dimension and family composition by age. Cluster 1 corresponds to larger families (average dimension: 3.2 members) with younger family members (average age: 28) and the Cluster 2 corresponds to smaller families (average dimension: 2.8 members) with older family members (average age: 53). However, clusters are mainly formed based on age differences. Therefore, the differences are more reliable in terms of the *per capita* consumption than in terms of the weekly consumption.

Cluster 1 has an average per capita consumption of 129 litres/(person.day) and Cluster 2 has an average per capita consumption of 163 litres/(person.day). These results show that older people tend to have higher consumption, which can be explained by the fact that they stay at home more time during the day than younger people that go to work.

Table 5. Characteristics of the family composition by age and family dimension clusters.

Cluster	Number of households	Average age	Average family dimension (n. of family members)	Per capita consumption [litres/(person.day)]			Weekly consumption [litres/(household.week)]		
				Minimum	Average	Maximum	Minimum	Average	Maximum
Cluster 1	31	28	3.2	50	129	286	473	2958	9966
Cluster 2	12	53	2.8	128	163	259	894	3158	5437

4.5. Assessment of household water use efficiency

Fig.2 and Fig.3 present the efficiency level for the households included in the family dimension and family composition by age clusters. These results exemplify the household assessment process. Results on efficiency levels for the family dimension and property type clusters can be found elsewhere [11].

The household efficiency level is calculated according to equation (1) and compares the efficiency of each consumer regarding their peers in the cluster they belong to, by relating each household consumption to the clusters' minimum or average efficiency. This analysis considered weekly consumption (Fig.2) and *per capita* consumption (Fig.3).

$$H_E = 100 + \frac{Cl - C_{CD}}{C_{CD}} \times 100 \tag{1}$$

in which

H_E	Household efficiency level	[%]
Cl	Cluster average or minimum	[litres]
C_{CD}	Consumption based on consumers' data	[litres]

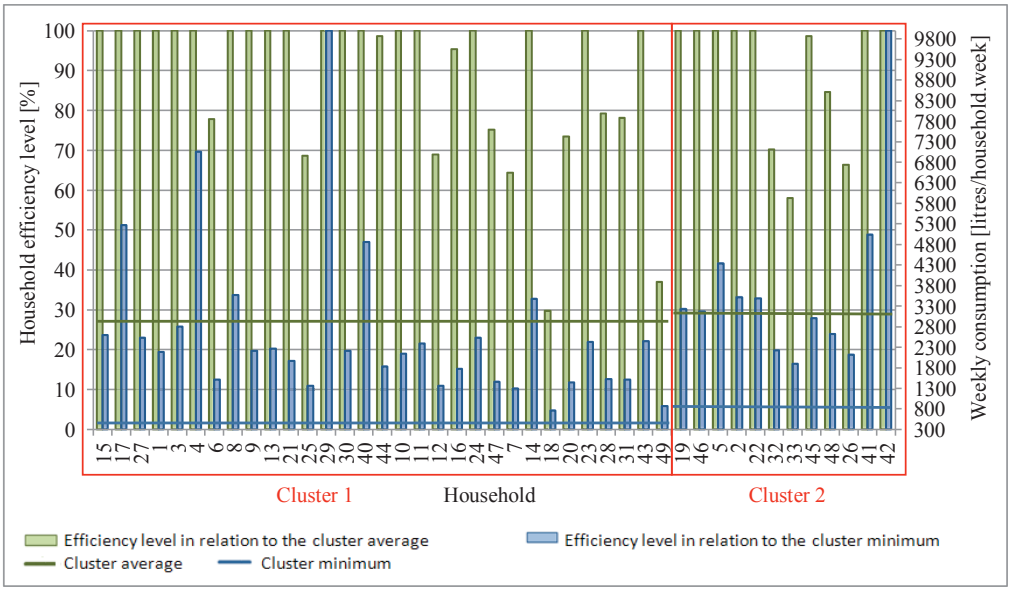

Fig.2. Family composition by age and family dimension clusters: weekly consumption analysis.

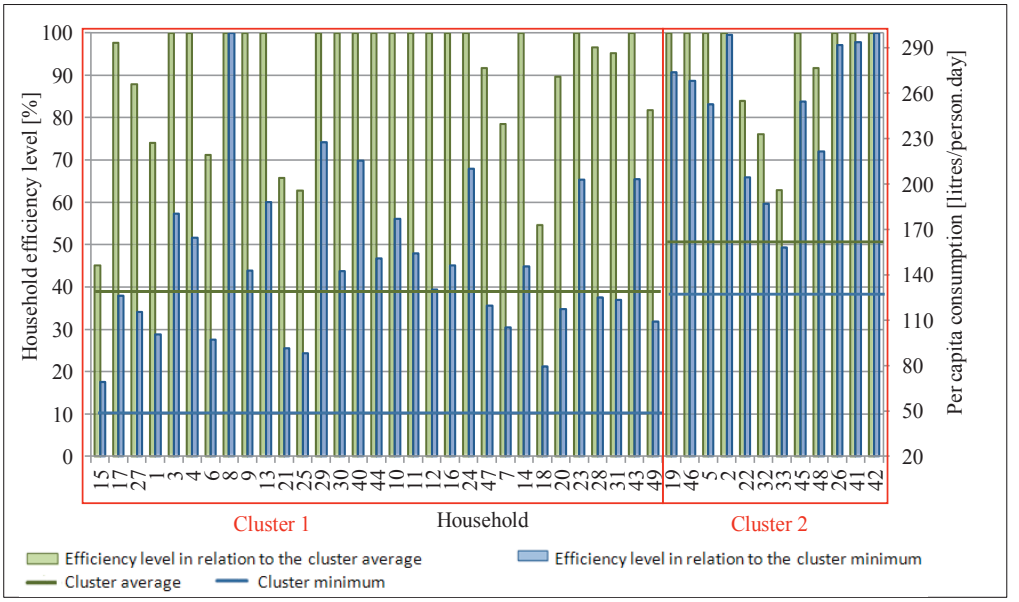

Fig.3. Family composition by age and family dimension clusters: per capita consumption analysis.

The x-axis in Fig.2 and Fig.3 represents the households within the family dimension and family composition by age cluster. The efficiency level for each household is also shown (left y-axis). The green bars represent the efficiency level (in %) of each household in relation to the average consumption of the respective cluster (right y-axis). The blue bars represent the efficiency level of each household in relation to the minimum consumption of the cluster (right y-axis).

In Fig.3, for instance, for household 15, the level of efficiency in relation to their cluster average consumption (129 litres/(person.day)) is 45%. This result means that household 15 should reduce their per capita consumption by 55% to become 100% efficient in relation to the cluster consumption average.

Fig.3, also shows that household 8 is the most efficient household of its cluster (Cluster 1), since its level of efficiency is 100%, in relation to both average and minimum consumption of Cluster 1. In Cluster 2, household 2 and household 42 are the most efficient households of the cluster.

Through this analysis, consumers can verify their deviation to the average and minimum of their respective cluster, and their performance when compared to other consumers with similar characteristics.

5. Conclusions

The main objective of this work is to present and to demonstrate a novel methodology to assess households' water use efficiency of indoor domestic water uses based on peer comparison.

For the current case study, results show that the assessment of the household water use efficiency should be made by comparing consumers within peer groups of similar family dimension, family composition by age and property type characteristics.

The application of this methodology to the case study provided an assessment of consumers' efficiency in relation to the consumers with similar characteristics, highlighting the most and the least efficient consumers in the different clusters. However, it should be noted that, if the methodology is to be applied to other cases, a specific cluster analysis should be made as the clusters type and composition depend on the socio-demographic features of the studied sample.

This work has also proven that there is not a unique or "better way" to carry out the peer comparison with the purpose of assessing water use efficiency, since different conclusions may be drawn about the efficiency of each household, depending on the type and characteristics of the cluster where a certain household is included.

Nevertheless, for performing peer comparisons and if information is available, socio-demographic clusters should be preferred over other type of clusters since consumers' are more fairly evaluated.

Through the methodology proposed in this paper, consumers are motivated to engage in water conservation efforts and to change wasteful behaviours.

Acknowledgements

The authors would like to thank to all the volunteer consumers that participated in the study that provided the results presented in this paper.

References

[1] P. Vieira, M., Almeida, J. Baptista, R. Ribeiro, "Household water use: a Portuguese field study," *Water Science & Technology: Water Supply*, 7, 2007, 193-202.

[2] M. Loh, P. Coghlan, "Domestic water use study: Perth," *Water Corporation*, Perth, Western Australia, 2003.

[3] C. D. Beal, R. A. Stewart, and K. Fielding, "A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption," *Journal of Cleaner Production,* 2011.

[4] B. Eck, E. Arandia, M. Barry, J. Barateiro, J. Coelho, A. Hashmi, J. Hayes, P. Kossieris, D. Loureiro, A. Mamade. C. Pantazis, N. Taheri, P. Vieira, G. Ziazopoulos, D222-Final report on the working prototype for iWIDGET, 2014.

[5] C. Matos, C. A. Teixeira, R. Bento, J. Varajão, and I. Bentes, "An exploratory study on the influence of socio-demographic characteristics on water end uses inside buildings," *Science of The Total Environment*, 2014, pp. 466–474.

[6] R. M. Willis, R. A. Stewart, D. P. Giurco, M. R. Talebpour, A. Mousavinejad, "End use water consumption in households: impact of socio-demographic factors and efficient devices," Journal of Cleaner Production, 2013, pp.107-115.

[7] L. Pinheiro, "Socio-demographic analysis for characterization of domestic consumption in water distribution systems", MSc Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2008 (in Portuguese).

[8] H.M. Corbella, D.S. Pujol, "What lies behind domestic water use? A review essay on the drivers of domestic water consumption," Boletín de la Asociación de Geógrafos Españoles, 50, 2009, 297–314.

[9] R. A. Lyman, "Peak and off-peak residential water demand," *Water Resources Research*, 28, 1992, 2159-2167.

[10] H. March, J. Perarnau, D. Saurí, "Exploring the links between immigration, ageing and domestic water consumption: The case of the Metropolitan Area of Barcelona," *Regional Studies*, 46, 2010, 229-244.

[11] C. Jorge, "Efficiency assessment of the household water use", MSc Thesis, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, 2014.

[12] S.-J. Hong, S. Y. Sohn, "Peer group analysis for introducing weather derivatives for a city," *Expert Systems with Applications*, 40, 2013, 5680-5687.