Water Resour Manage (2013) 27:2155–2177
DOI 10.1007/s11269-013-0281-8

An Approach to Disaggregating Total Household Water Consumption into Major End-Uses

Sara Fontdecaba · José A. Sánchez-Espigares · Lluís Marco-Almagro · Xavier Tort-Martorell · Francesc Cabrespina · Jordi Zubelzu

Received: 26 September 2012 / Accepted: 16 January 2013 /
Published online: 27 January 2013

Abstract The aim of this project is to assign domestic water consumption to different devices based on the information provided by the water meter. We monitored a sample of Barcelona and Murcia with flow switches that recorded when a particular device was in use. In addition, the water meter readings were recorded every 5 and 1 s, respectively, in Barcelona and Murcia. The initial work used Barcelona data, and the method was later verified and adjusted with the Murcia data. The proposed method employs an algorithm that characterizes the water consumption of each device, using Barcelona to establish the initial parameters which, afterwards, provide information for adjusting the parameters of each household studied. Once the parameters have been adjusted, the algorithm assigns the consumption to each device. The efficacy of the assignation process is summarized in terms of: sensitivity and specificity. The algorithm provides a correct identification rate of between 70 % and 80 %; sometimes even higher, depending on how well the chosen parameters reflect household consumption patterns. Considering the high variability of the patterns and the fact that use is characterized by only the aggregate consumption that the water meter provides, the results are quite satisfactory.

Keywords Water pattern recognition · Use of water · Domestic consumption · Water profile · Water disaggregation

1 Introduction

Many efforts have been made to understand the factors and behaviors that cause high variations in per capita water consumption among various regions, cities and people. With

S. Fontdecaba · J. A. Sánchez-Espigares · L. Marco-Almagro · X. Tort-Martorell (✉)
Department of Statistics and Operational Research,
Universitat Politècnica de Catalunya (UPC) – Barcelona Tech, Barcelona, Spain
e-mail: xavier.tort@upc.edu

F. Cabrespina · J. Zubelzu
Aigües de Barcelona, AGBAR Group, Barcelona, Spain

this in mind, Aigües de Barcelona conducted a study (Fontdecaba et al. 2012) to compare the consumption of clients within homogeneous socioeconomic groups. One significant application of this project was higher accuracy in forecasting future water demand. Realizing that a relationship exists between socioeconomic characteristics and water consumption habits was a step forward because, in general, water companies know very little about the patterns and ways their customers use water. Obviously, knowing how clients use the water would be even better as it would allow the companies to, for example, quantify the impact of water efficiency measures or new policy initiatives, as well as facilitating more accurate forecasts. Thus, it was logical (and challenging) for Aigües de Barcelona to seek this information by researching how to disaggregate the whole domestic water consumption into the different major end-uses.

The problem is difficult because individual metering of water devices in homes is awkward and expensive. Further, the absence of pertinent data makes it unfeasible to identify and quantify water consumption by devices with the same accuracy that has been achieved for gas (Yamagami and Nakamura 1996) and electricity (Farinaccio and Zmeureanu 1999) consumption.

The objective of the work presented here is to show that whole-house meter data can provide quite accurate end-use water consumption data. The project was conducted on behalf of and with full support from Aigües de Barcelona, which was interested in learning about their clients' consumption habits. The first stage of the project was the analysis of water consumption patterns in the city of Barcelona by monitoring device consumption in different homes from different socioeconomic segments. Two types of information were recorded: the water consumption from the whole-house water meter (read every 5 s) and synchronized information from the devices consuming the water. Thus, the monitored group provided baseline information for tackling the problem of identifying the devices by using only data from the meter. The problem was difficult because of the scarcity of the information on which to base the assignation and also because consumption habits varies highly among people, households and even individuals within the same household.

This paper is organized in the following manner: Section 2 briefly reviews the scope and data used to carry out the study in the area of Barcelona; Section 3 provides the most relevant results of the data analysis; Section 4 explains the methodology followed in uses recognition patterns; Section 5 highlights and discusses the most relevant results obtained in the Barcelona case; and finally, Section 6 presents some conclusions.

2 Case Study and Data Used

2.1 Data Sources

In order to ensure that the sample structure was as robust as possible, the property sample structure was designed by taking into account all major factors that were known to influence household water consumption. In the Barcelona area, clients can be grouped into segments (six segments) with homogeneous consumer habits, which are known to be related to socioeconomic status (Fontdecaba et al. 2012). That was the reason behind considering the variability between households as an important source to consider when designing the sample; accordingly, we wanted to guarantee that households from all 6 segments were included. These segments are distributed on the Barcelona map in a rather predictable pattern; so to avoid geographical bias, the location of the eight households used in this study was also taken into account.

To monitor each of the households, we designed a system which monitored two complementary blocks of data: meter traffic flow and each of the different water consumption points inside the household (kitchen basins, cistern, washing machine, etc.).[1] The data collected was consolidated and stored on a remote server for later analysis.

The first source of data (at the household level) was not difficult to obtain, since it is already collected by the metering system of Aigües de Barcelona. Monitoring water use within the household was the main problem, and we opted for an integrated solution which involved placing a set of non-intrusive flow sensors (called flow switches) at different points on the pipe section leading to a single consumption.

The sample frequency and the water consumption level of detail were key questions regarding the assignation, and they were decided according to the main objective of the project. Thus, we considered the minimum time period that the devices could be in use and could be logged. In the end, the two sources of data were collected every 5 s over a period of 3 months, between December 2009 and February 2010.

2.2 Data Management

Information from the meter was collected every 5 s with a 0.1 liter resolution, using a concentrator that sent periodic requests to the electronic meter, which registered, dated and saved the totalized flow and its characteristic attributes. If no flow variation occurred when compared to the previous request, it was ruled out.

The flow switches were permanently aware and attached directly to wireless terminals that, every 5 s, reported to a network master whether water passed through (1) or not (0). The master acted as a coordinator at the household level and also as a gateway to an on-site storage facility that compiled and registered all received data via GPRS to the remote server. This approach allowed the data to be received every day through GPRS and avoided the use of data loggers, which would have been more complex and expensive.

To sum up briefly, the monitoring process in the electronic meter gave us progressive information about the meter counter and the flow over time. The flow switches provided information about where and when water was used in the household (washing machine, shower, kitchen basin, etc.). Considering the two together offers a picture of household consumption, such as in Fig. 1.

The system provided us with a data base containing the water use and meter data of 8 households over a period of 3 months; a huge amount of data. To automatically assemble, elaborate on and consolidate such information was a meticulous process. To illustrate this point, consider the spreadsheet snapshot below, which corresponds to a particular household in which the test was conducted (Fig. 2).

Number 1 indicates the rows that were sorted according to the time line evolution, with every row corresponding to a sample. The usual timing from sample to sample is 5 s (monitoring period). Number 2 (Column B) shows the time, with the day, month and hour when the sample was captured. Number 3 (Column C) corresponds to the meter's totalized reading at the time of the sample. Number 4 (Column D) is the flow in liters per second coming into the household. As a visual aid, the color code is proportional to the flow intensity. The group of columns represented by number 5 is the water use where each column corresponds to a unique, recognizable consumption point. The 0/1 sequence found

[1] From now on we will use the word "device" to refer to the different things causing consumption weather they are appliances, showers, taps or cisterns.

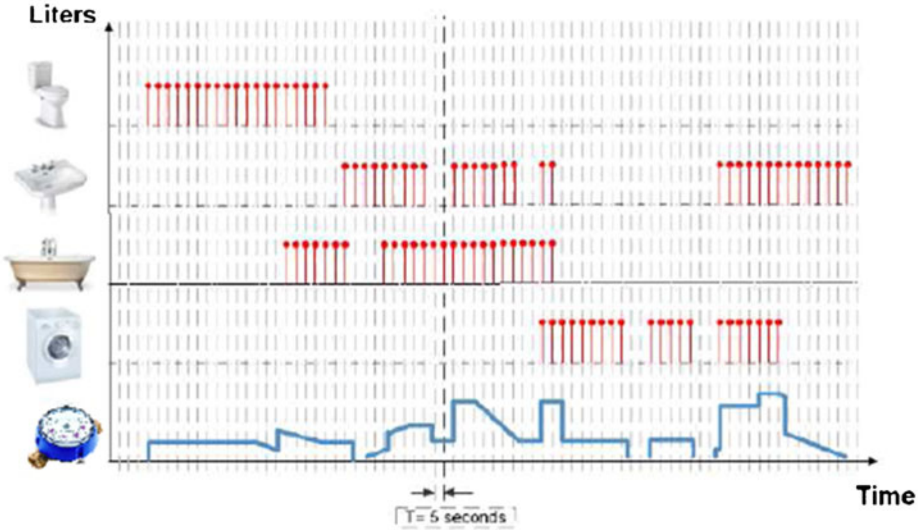


Fig. 1 Household consumption data

in any water use column indicates whether or not the particular consumption point was used. The red and white backgrounds provide visual aid.

Despite the fact that two flow-switches were installed for hot and cold water, our interest was in detecting use activation without distinguishing between the two. For this reason, the sequences of 0/1 corresponding to the two types of water were combined.

Considering all the 8 households, a total of 12 different cisterns, 8 washing machines, 8 kitchen basins, 8 bath taps, 5 dishwashers and 10 showers are in the sample.

3 Data Analysis

In order to investigate and understand the variability in microcomponent data (data consumptions disaggregated at the devices level), the dataset was analyzed using a number of statistical methods. The main characteristics analyzed were the frequency of the water use (understood as the number of times each device is used in each property per day) and the volume per use.

Even after measures were taken to "guarantee" the data quality of the 8 monitored households and several corrections and repairs were performed on the loggers during the monitoring period, the database contained mistakes and inconsistencies. After thorough cleaning, the database was ready for analysis. It is worth mentioning that several of the problems were related with the loggers from 3 households that had dishwashers; thus, the information on the consumption patterns of this appliance is scarce. However, as we will see in the developed procedure, it is sufficient.

3.1 Definitions and General Characteristics

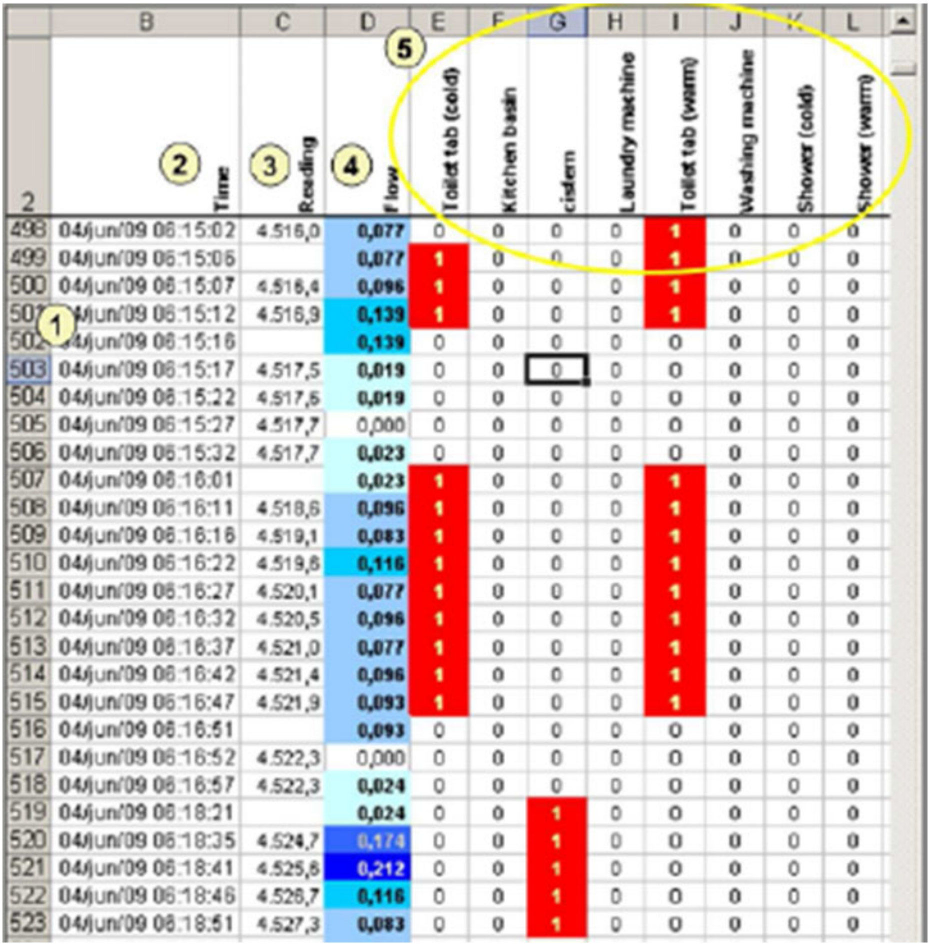
Before beginning the statistical analysis to characterize the consumption patterns of each device, it is necessary to establish some specific definitions.

Water use continuous period of time in which the same device takes water. It is identified in the database by using a combination of flow and time criteria (Fig. 3).

Some devices, such as the cisterns, consume water in a continuous way. When they start, they begin consuming water and, when they end, consumption ends. However, dishwashers and washing machines take water in a discontinuous way. In general they take some water at the beginning, then stop the intake and perform other operations. A few minutes later, they take in water again. This pattern is repeated at different intervals until the device discontinues its use. In other words, their total water consumption is composed of different use groups. In these cases it will be necessary to define the concept of program.

Program Consecutive groups of water uses belonging to the same device that define the full use of the device. It can be composed of one or more water uses (Fig. 3).

The consumption of other devices, such as the shower, can be characterized by a program with a single water use or a program with several water uses. It depends on the shower habits of the individual.



	B	C	D	E	F	G	H	I	J	K	L
2	Time	Reading	Flow	Toilet tab (cold)	Kitchen basin	cistern	Laundry machine	Toilet tab (warm)	Washing machine	Shower (cold)	Shower (warm)
498	04/jun/09 06:15:02	4.516,0	0,077	0	0	0	0	1	0	0	0
499	04/jun/09 06:15:06		0,077	1	0	0	0	1	0	0	0
500	04/jun/09 06:15:07	4.516,4	0,096	1	0	0	0	1	0	0	0
501	jun/09 06:15:12	4.516,9	0,139	1	0	0	0	1	0	0	0
502	4/jun/09 06:15:16		0,139	0	0	0	0	0	0	0	0
503	04/jun/09 06:15:17	4.517,5	0,019	0	0	0	0	0	0	0	0
504	04/jun/09 06:15:22	4.517,6	0,019	0	0	0	0	0	0	0	0
505	04/jun/09 06:15:27	4.517,7	0,000	0	0	0	0	0	0	0	0
506	04/jun/09 06:15:32	4.517,7	0,023	0	0	0	0	0	0	0	0
507	04/jun/09 06:16:01		0,023	1	0	0	0	1	0	0	0
508	04/jun/09 06:16:11	4.518,6	0,096	1	0	0	0	1	0	0	0
509	04/jun/09 06:16:16	4.519,1	0,083	1	0	0	0	1	0	0	0
510	04/jun/09 06:16:22	4.519,6	0,116	1	0	0	0	1	0	0	0
511	04/jun/09 06:16:27	4.520,1	0,077	1	0	0	0	1	0	0	0
512	04/jun/09 06:16:32	4.520,5	0,096	1	0	0	0	1	0	0	0
513	04/jun/09 06:16:37	4.521,0	0,077	1	0	0	0	1	0	0	0
514	04/jun/09 06:16:42	4.521,4	0,096	1	0	0	0	1	0	0	0
515	04/jun/09 06:16:47	4.521,9	0,093	1	0	0	0	1	0	0	0
516	04/jun/09 06:16:51		0,093	0	0	0	0	0	0	0	0
517	04/jun/09 06:16:52	4.522,3	0,000	0	0	0	0	0	0	0	0
518	04/jun/09 06:16:57	4.522,3	0,024	0	0	0	0	0	0	0	0
519	04/jun/09 06:18:21		0,024	0	0	1	0	0	0	0	0
520	04/jun/09 06:18:35	4.524,7	0,174	0	0	1	0	0	0	0	0
521	04/jun/09 06:18:41	4.525,6	0,212	0	0	1	0	0	0	0	0
522	04/jun/09 06:18:46	4.526,7	0,116	0	0	1	0	0	0	0	0
523	04/jun/09 06:18:51	4.527,3	0,083	0	0	1	0	0	0	0	0

Fig. 2 Water use and meter data collection

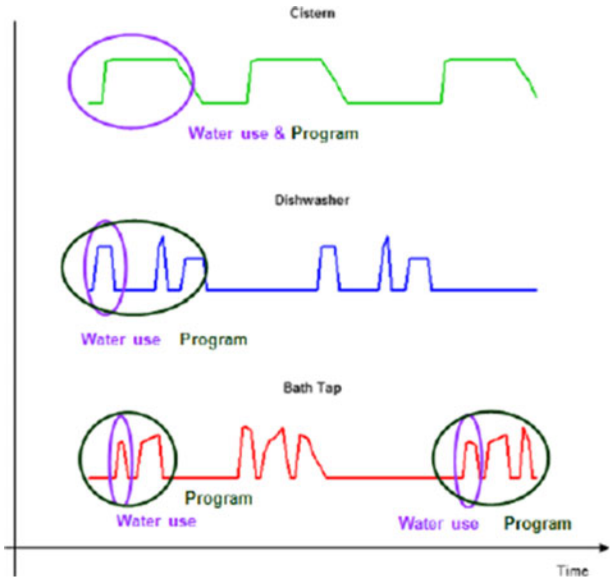


Fig. 3 Definition of water use and programs

The kitchen basin and bath tap are very complicated (in many cases impossible) for distinguishing whether different consecutive water uses belong to the same program or to a succession of single-use programs. Different households may have very different patterns of tap usage.

Given the need to correctly identify the programs, it was necessary to establish criteria for considering the limits between programs. The frontier between programs was based on the time period between two water uses by the same device; when this period exceeded a specified time—which was different for each type of device—the consumption was assigned to two different programs. Table 1 shows the maximum time period allowed between continuous water uses which were considered to be in the same program:

Both programs and water uses were statistically analyzed in order to characterize the water consumption behavior of each device. For example, in the case of the washing machines, it would then be possible to answer questions like these: How long is a washing machine program? How many times does the washing machine take in water? For how long does the washing machine take in water?

Table 1 Maximum time (in min.) allowed

Device	Minutes
Cistern	0
Washing Machine	60
Dishwasher	60
Shower	15
Kitchen Basin	10
Bath Tap	10

3.2 Descriptive Analysis of Programs and Water Uses

One of the main objectives of the descriptive analysis is to understand the consumption profile of different devices.

In order to have a robust descriptive analysis—one that would give the best possible unbiased information on how each device consumes water—only programs that began and finished without any interference from other devices were taken into account (67 % of the total number of programs). For the 8 monitored properties, Table 2 shows the total number of available programs and water uses.

The total daily household consumption is nearly 93 l per house. To interpret this number, it is necessary to keep in mind that not all devices in the households were monitored and that the results only include the devices when they are running without simultaneities. The shower represents 40 % of this daily consumption and the cistern 32 %. The kitchen basin and the bath tap constitute 21 %, and a small percentage of about 6 % is for the washing machine and the dishwasher.

Within each property, the distribution of the percentages for the total consumption of water was very similar. The higher percentages covered showers and cisterns. Bath and kitchen taps constitute the second block of consumption to be considered, followed by washing machines and dishwashers.

The percentage of total water consumption covered by devices does not differ from results reported by the Statistical Office of the European Commission (EUROSTAT 2007) and several other published works (Mayer et al. 2003 and Richter and Stamminger 2012).

3.3 Analysis of Possible Influential Factors in Domestic Water Consumption

Considering the main objective of distinguishing the consumption profiles of each of the devices, it would be useful to find variables that would help differentiate them. The pace of life today runs on timetables, therefore it is reasonable to assume that the time of day or day of the week (i.e. work days vs. weekends) can influence consumption habits.

3.3.1 Time of Day

Daily water consumption per hour, as averaged across all 8 households, is not useful in characterizing use by any of the devices. All the programs are more or less distributed

Table 2 Statistics of program frequency and volume for each device

Device	Number of water uses	Number of programs	Frequency of use (prog./property/day)	Consumption per prog. (liters/prog.)	Household Consumption (liters/property/day)	% of total Consumption
Cistern	3.077	3.077	6,88	5,56	30,10	32,11
Washing Machine	575	69	0,33	38,06	5,62	4,92
Dishwasher	538	88	***	***	***	***
Shower	1.261	591	1,31	36,37	35,47	40,31
Kitchen Basin	8.072	3.068	6,90	2,69	14,20	15,47
Bath Tap	3.362	2.183	6,80	1,35	4,75	5,52

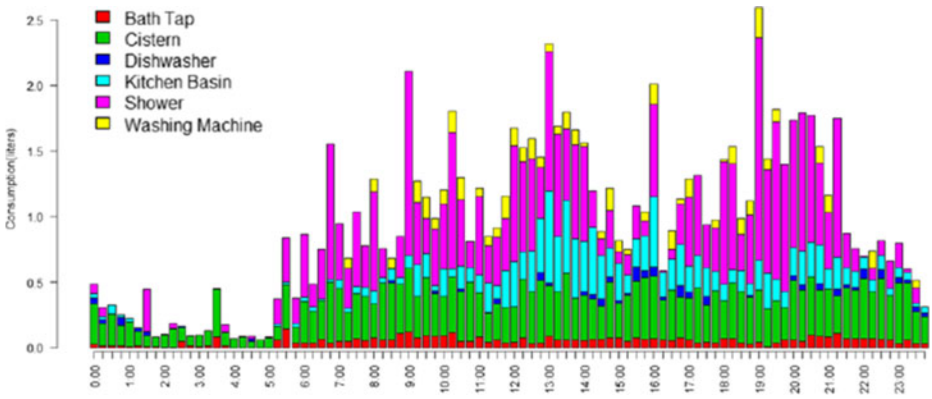


Fig. 4 Average water consumption per day and hour

throughout the day without any significant considerations. Cisterns and showers are distributed throughout the whole day, while kitchen basins register their maximum use at lunch time. Washing machines are distributed from 7.00 to 23.00, and dishwashers normally appear in the early afternoon (14.00–16.00). During the morning (6.00–10.00), cisterns and showers constitute the main use (Fig. 4).

3.3.2 Work Days and Weekend Water Consumption

People do not consume water differently on work days and weekends. Table 3 shows the distribution of the programs used during work days and weekends across the 8 properties (71.9 % of programs are executed during work days and 28.1 % during weekends). All weekday holidays were considered as labor days. The small differences are insignificant to the purpose of this study for detecting device consumption (Table 3). The distribution has also been tested within every property, and none of them show significant differences between the types of days considered.

3.4 Descriptive Analysis of the Devices

In order to identify which device has consumed the water recorded by the meter, it is necessary to characterize (model) what kind of consumption each device produces. The level of detail in the microcomponent dataset used in this study reasonably represents the

Table 3 Distribution of the program during weekdays, weekends and in general

Device	Labor	Weekends	General
Cistern	35.7 %	33.2 %	33.9 %
Washing Machine	0.8 %	0.8 %	0.8 %
Dishwasher	1.1 %	0.9 %	1.0 %
Shower	6 %	6.7 %	6.5 %
Kitchen Basin	33.2 %	34 %	33.8 %
Bath Tap	23.3 %	24.4 %	24.1 %

Table 4 Descriptive statistics of the cisterns

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration (sec.)	4,30	29,40	40,60	43,24	55,30	115,80	17,97
Consumption (l.)	1,00	3,49	4,62	5,56	7,68	21,60	2,86

programs for characterizing and analyzing the devices. The descriptive analysis that follows will help develop the procedure for identifying microcomponents based on the information provided by the water meter.

3.4.1 The Cistern

Cisterns are the most used device in domestic habitats. Their total duration and the minimum and maximum levels they consume are the most significant results for characterizing their programs. Table 4 shows the descriptive statistics averaged across all the 8 properties (Min.: minimum, 1st Qu.: first quartile, 3rd Qu: third quartile, Max: maximum and Stdev: standard deviation).

According to the descriptive analysis, cisterns seem to take in nearly 5.5 l of water in 43 s. These are the combined results of both dual- and single-flush cisterns, and therefore the cause of such a high standard of deviation. In the sample, 40 % of the cisterns were water efficient, dual-flush toilets with reduced flush and time consumption.

3.4.2 The Shower

Showers (and baths) are among the devices which consume the most water. Their consumption basically depends on flow rates and the duration in which they are turned on. Depending on individual habits and preferences, people may turn the water off periodically, decrease the water pressure at certain times, or they may have shorter showers with a constant flow of water.

The statistics used to characterize all of these habits are (per shower): duration, consumption, and the number of water uses (remember that this represents the number of times that the shower tap is turned on and off during a program, i.e., a single shower). The results are summarized in Table 5.

3.4.3 The Washing Machine

There are a wide range of washing machines on the market. They often have as many as ten programs for washing, rinsing and spinning. Moreover, energy-efficient washing machines

Table 5 Descriptive statistics of the showers

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration (time)	0 min. 5 s.	1 min. 6 s.	4 min. 40 s.	5 min. 20 s.	7 min. 40 s.	21 min. 3 s.	4 min. 22 s.
Consumption (l.)	1,00	8,35	33,2	36,37	52,41	240,60	31,18
Water uses	1,00	1,00	2,00	2,13	3,00	14,00	4,2

use sensors to detect the size of a load and accordingly determine the appropriate amount of water (some even incorporate fuzzy logic strategies to vary the amount of water while running). All these factors increase the variability among programs and hinder their characterization.

The descriptive statistics are based on washing machine programs separated from any other device program. The useful variables for characterizing washing machines are: consumption, the number of uses during the programs, the duration of the program and the duration of time that the washing machine takes in water. Figure 5 shows the differences between the two last statistics.

The main summary statistics are presented in Table 6:

3.4.4 The Dishwasher

Dishwashers consume water very similarly to washing machines. They have different programs and they are able to use the appropriate amount of water, depending on the load; so variability is very high. Therefore the variables considered are the same, and the summary statistics are presented in Table 7.

The numbers in Table 7 have to be considered very cautiously, because they are based on problematic data (loggers not functioning properly). However, and given that we had very few "clean" dishwasher observations, some facts could be derived, especially those corresponding to the columns in bold (numbers in the other columns have been represented in smaller typeface to consider their unreliability.

3.4.5 The Internal Taps

Tap use on a property is involved in various activities and housework. This is the reason behind the very high variability.

Table 8 shows the main statistics for all the programs associated to the kitchen basin:

It is clear that there are extreme programs with an extreme maximum duration and consumption. Most of the programs have short duration and low consumption; but there are a negligible number of programs (again, for characterizing and identifying microcomponents) with very high consumption as well as very long duration.

The results of calculating the same statistics for the programs belonging to the microcomponents of the bath tap are presented in Table 9.

Once again eliminating the extreme programs, the distribution of duration and consumption is close to the distributions of the kitchen basin. This feature makes it difficult to

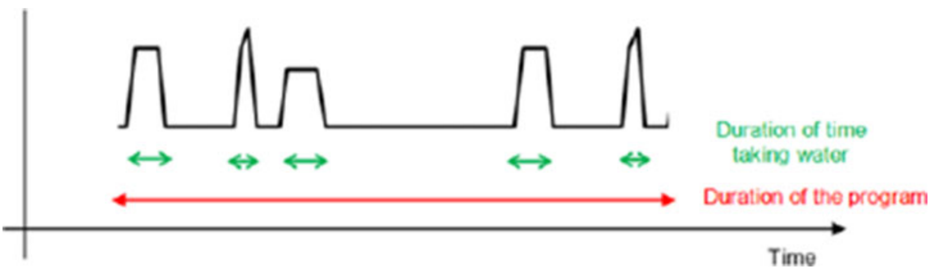


Fig. 5 Differences between duration of program and duration of time taking in water

Table 6 Descriptive statistics of washing machines

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration Program (time)	8 min. 15 s.	33 min. 55 s.	40 min. 45 s.	47 min. 33 s.	57 min. 50 s.	122 min 35 s.	24 min. 33 s.
Duration Taking in water (time)	1 min. 25 s.	3 min. 50 s.	5 min. 30 s.	5 min. 10 s.	6 min. 2 s.	11 min. 28 s.	2 min. 15 s.
Consumption (l.)	1,15	30,05	42,55	38,06	47,34	61,69	12,75
Water uses	3,00	7,00	8,00	8,31	10,00	14,00	2,04

Table 7 Descriptive statistics of dishwasher programs

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration Program (time)	0 min. 10 s.	17 min. 12 s.	32 min. 55 s.	45 min. 56 s.	**84 min. 32 s.**	**135 min 10 s.**	40 min. 02 s.
Duration Taking in water (time)	0 min. 5 s.	2 min. 10 s.	2 min. 44 s.	2 min. 32 s.	**3 min. 14 s.**	**4 min. 42 s.**	1 min. 13 s.
Consumption (l.)	0,3	5,27	10,79	10,16	**13,3**	**28,5**	7,03
Water uses	1,00	2,00	5,00	6,11	**9,00**	**14,00**	4,2

Table 8 Descriptive statistics of the Kitchen Basin

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration (sec.)	0,9	5,20	14,70	38,96	37,62	1.087,0	72,44
Consumption (l.)	0,02	0,37	0,9	2,68	2,62	102,6	5,33
Water uses	1,0	1,0	1,0	2,6	3,0	31,0	3,02

differentiate the processes of bath taps and kitchen basin programs. Clearly it is very complicated, if not impossible, to find a criterion to separate and distinguish them.

Since consumption differences between kitchen basins and bath taps are not easy to identify, from now on the two devices will be treated under the unique use of internal taps.

3.5 Variability in the Database

A very interesting finding appeared during the analysis of device consumption and duration patterns inside each household. Our initial expectation was that variability would be much lower; that is, that the variability shown in the above analysis would vary mainly "between households". To our surprise, the independent data analysis of each household showed a similar degree of variability. In other words, the "within household" variability is at least as significant to total variability as that of the "between household" differences.

Figure 6 shows that the washing machine in household number 8 (Fig. 6a) exhibits different consumption (l.) and duration (sec.) profiles. It is very common that modern washing machines allow programming of various features such as temperature or rotation speed settings, among others. For this reason, the type of clothing in a household can cause great variation in how a washing machine uses water. In comparison to household number 2 (Fig. 6b), we can see that the same program is normally used for all loads; thus, there is low variability within the household. The use of a single program differs from household 8, and therefore there is greater variability between households. Household numbers 8 and 2 represent the respective maximum and minimum intra-variability among washing machine programs.

A similar effect occurs with the cistern profiles shown in Fig. 7. In this case, there is high variability within household number 8, where we can see (Fig. 7a) differences in cistern duration and consumption, probably because both a single and a dual flush cistern are present in the same household. Otherwise, household number 5 (Fig. 7b) is dominated by only one single-flush cistern, with a 10-liter mean consumption and 50-second duration. Household numbers 8 and 5 represent the respective maximum and minimum intra-variability cistern profiles.

Table 9 Descriptive statistics of the Bath Tap

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stdev.
Duration (sec.)	3,50	5,20	14,70	22,51	29,40	1753,4	44,59
Consumption (l.)	0,1	0,32	0,72	1,34	1,70	22,97	1,79
Water uses	1,0	1,0	1,0	1,54	2,0	15,0	1,17

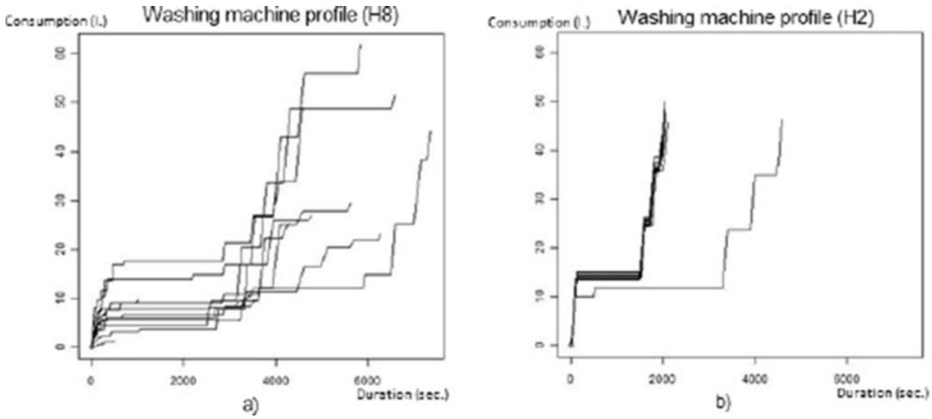


Fig. 6 Example of washing machine variability between household 8 (H8) and household 2 (H2), and within them

4 Methodology

Statistically speaking, there was no previous, specific methodology for identifying which device consumes the water that passes through the water meter. Therefore, we faced the difficult task of devising a methodology. The proposed approach is a combination of a set of statistical techniques. The final procedure is the result of many trial and error efforts, and only a few of them are presented here.

4.1 First Approach: Profile Recognition

Given the project objectives, we first tried using pattern recognition techniques to identify which devices consumed the water indicated by the meter. Pattern recognition aims to classify different patterns based either on a priori knowledge or on statistical information extracted from the patterns. It seemed to be a good strategy, since each device has a pattern (or profile) based on the duration and the level of consumption. The first step tried to define

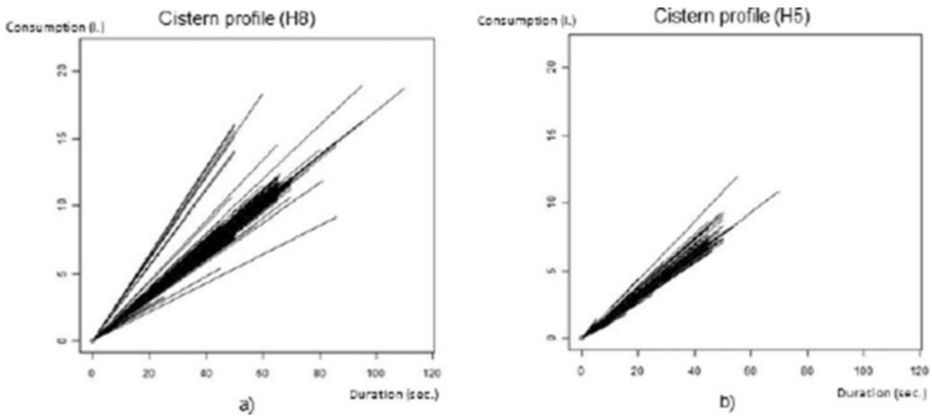


Fig. 7 Example of cistern variability between household 8 (H8) and household 5 (H5), and within them

the patterns to be classified. In this case the patterns to be classified were groups of water measurements that defined different device profiles. The schematic idea is to compare the pattern of a given device over the actual consumption profile (the measured values) and to do this over time; every time a god fit is identified this consumption is labeled as corresponding to this device and the corresponding consumption is extracted. The process is repeated for all device patterns. However, we tried several approaches, all of which were effective only for the cistern.

For all the remaining devices, the curve could not be estimated (or would be estimated with a very high variability, rendering it useless) because of the high variability from position and duration of water use inside each program. Moreover, the simultaneity of programs had a direct impact on flow volume, which distorted the metered value. Thus, the proposal of pattern recognition had to be discarded.

4.2 The Importance of Water Uses

At this point of the process, studying the water uses—i.e., the location and duration within the programs—would solve the variability problem.

Several attempts failed to fit a curve (estimated by several different procedures) to the water uses, though they took into account the program and considered them independent entities. The water uses also varied greatly, but they also indicated a good possibility of finding a statistical solution for summarizing the information acquired.

The proposal was to simplify the water use into its primary statistics. That is, using duration (in seconds), consumption (in liters) and flow (in liters/second) as working variables instead of the curve. This proposal simplified the information and allowed better treatment of meter variability (Fig. 8).

4.3 Combination of Water Uses

Once the water uses considered as independent entities (not part of programs) were characterized, the next problem was to associate them to a program based only on the information

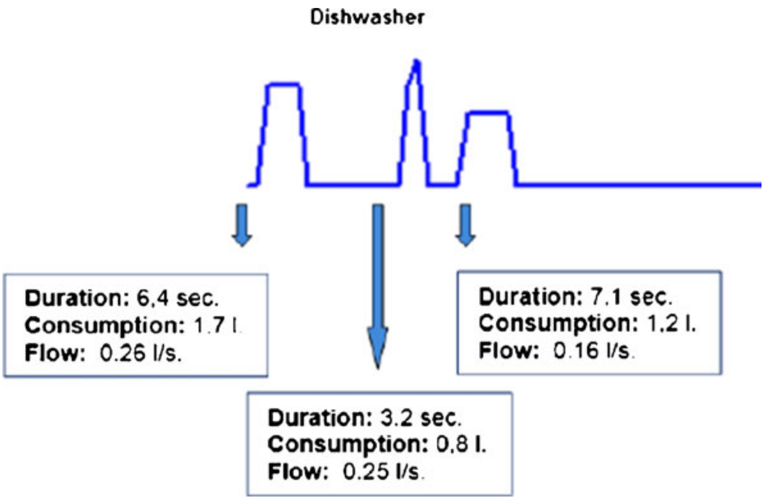


Fig. 8 From curve characterization to water use characterization

provided by the water meter. The idea was to find combinations of water uses that constituted a feasible program. Of course, on many occasions there were many different ways to combine water uses that produced feasible programs; thus a new problem arose: choosing among them (Fig. 9). The selection criteria was very important and it was necessary to keep in mind that most devices (washing machines, dishwashers, showers and internal taps) use water more than once during a program. The one exception is cisterns.

4.4 Evaluation of Variants in Identifying Water Uses

The combinatorial problem of grouping a random number of water uses associated to programs is very extensive. Computationally, not all combinations could be tested. Thus, it was necessary to create a methodology based on three steps:

1. Design a criterion for assigning the combinations.
2. Create a mechanism for producing variants that increase the probability of testing the correct combination.
3. Evaluate the programs and assign the use.

4.4.1 Designing a Criterion for Assigning the Combinations

Common sense indicates that a program is composed of a logical number of nearby water uses. We would consider that two consecutive water uses will belong to different programs when the time period between them exceeds 60 min for the washing machine and

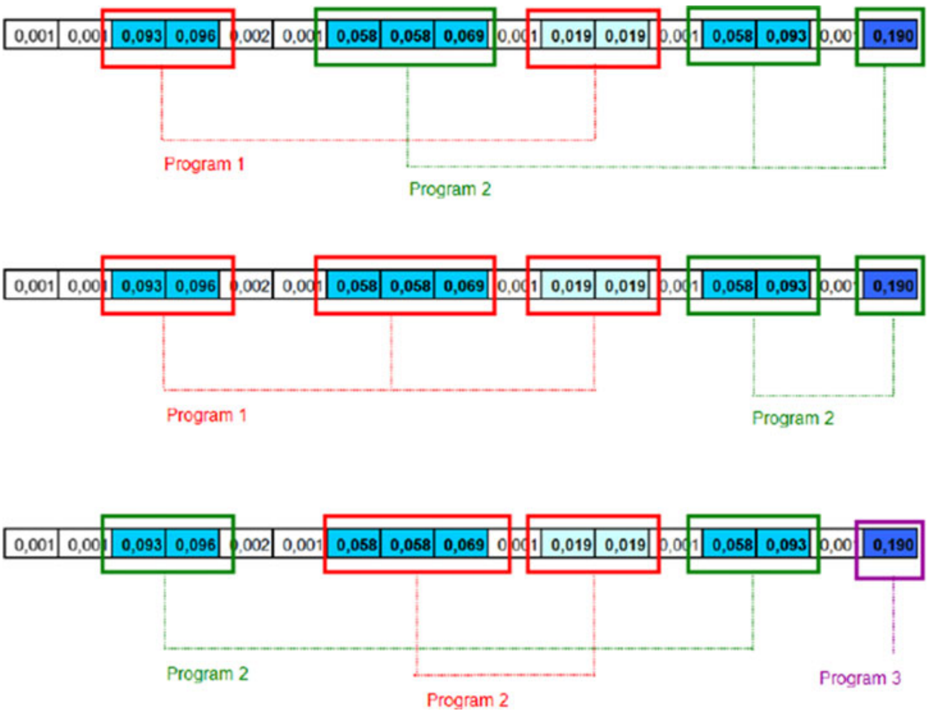


Fig. 9 Multiple combination of water uses to create programs

dishwasher, 15 min for the shower and 10 min for internal taps. Cisterns are considered to be generated by a singular water use (Fig. 10).

Because of the discontinuous water uses belonging to the same programs, especially the washing machines and dishwashers, it was necessary to identify some criteria for differentiating devices. In describing the different events, it was evident that the initial and final water uses of each device were quite particular, which helped classify the total program. This was the reason why the initial and final uses were used to evaluate the potential variants.

4.4.2 Generating the Variants

The first idea was to generate different possible programs (called "variants") by randomly grouping consecutive water uses. However, after some trials the procedure presented several complications, one of which was the improbability of calculating the correct program; so that procedure was discarded.

Then the investigation proceeded toward applying non-random heuristics. After designing and evaluating several heuristics, one was chosen which provided a very good relationship between computing time and accurate program identification. It began with a "compartment" composed of a consecutive number of water uses. This number had to be large enough to contain any possible program and small enough to maintain software

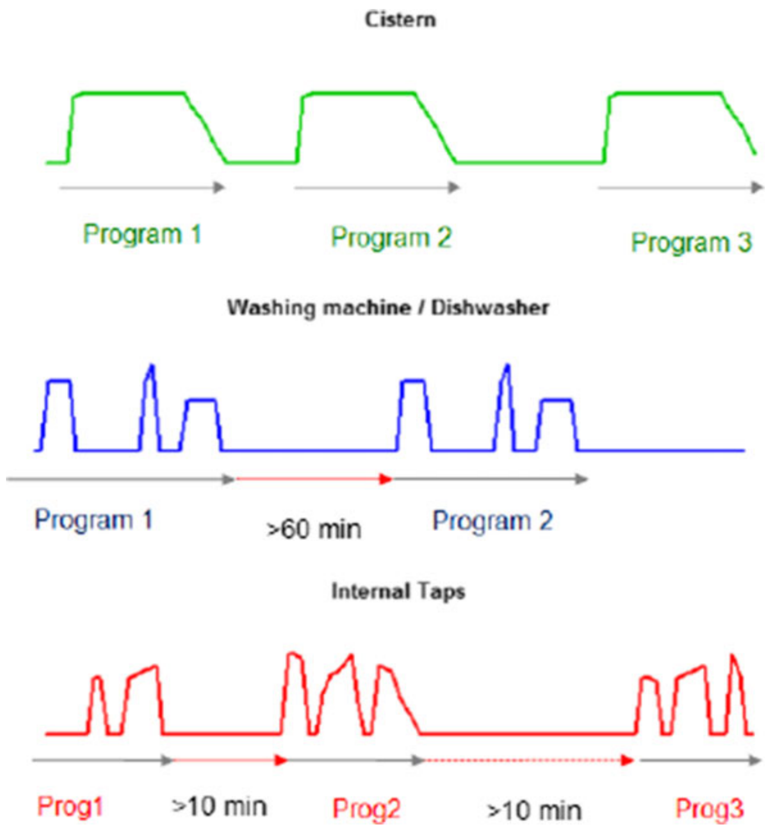


Fig. 10 Example of maximum time allowed between water uses within the same program

execution time within feasible levels. The size of this compartment was treated as a random variable, so it was optimized to better identify the microcomponent within a reasonable time limit.

Once a compartment is established (keep in mind that it is moved along the water meter reading time series), the algorithm creates all possible variants obtained by removing one, two or three combinations of water uses inside it. Then each water use removed is evaluated to be classified as a possible cistern. All the variants generated in this way are tested as possible programs of any of the devices. In general, considering a compartment of *m* water uses, the number of variations of *k* water uses is:

$$Var_{m,k} = \frac{m!}{(m-k)!k!}, \quad where\ m! = m \cdot (m-1) \cdot (m-2) \cdot \ldots \cdot 1$$

4.4.3 Evaluation of Variants for Identifying Programs

All the variants obtained with this heuristic are sequences of water uses which potentially belonged to any of the available microcomponents. To evaluate and classify each of the device programs, the heuristic calculated five variables for each sequence. These statistics were:

Referring to the program:

X_1: Total program duration (sec.): time between the beginning of the initial water use and the end of the final water use.
X_2: Total consumption (l.): total amount of water
X_3: Time of service (sec.): total time water is flowing.
X_4: Mean Flow (l/sec.): total consumption divided by time of service.
X_5: Number of water uses.

Referring to the initial water use:

X_6: Total consumption (l.): total amount of water that flows during the initial water use
X_7: Time of service (sec.): duration of initial water use
X_8: First Delay (sec.): time between the end of the initial water use and the beginning of the second (by convention, 1 s if there is only one water use)

Referring to the final water use:

X_9: Total consumption (l.): total amount of water that flows during the final water use
X_{10}: Time of service (sec.): duration of the final water use.

Having the statistics, the program uses the likelihood function to apply a mathematical methodology related to the matching templates. In this problem the templates are the descriptive analysis of each device explained in section 3.4 (Descriptive analyses of the devices). The best approximation to the cistern model is the median values of duration, consumption, etc., shown in the statistical summary of the eight monitored properties. Thus, there are 5 available templates (or models), each one related to an device: cistern, shower, washing machine, dishwasher and internal taps. These templates are general and robust, since they summarize the behavior and variability of the different devices used by different users in different households.

The likelihood function is a function of the parameters of a statistical model. In this case, it depends on all the ten variables considered. In our case, the likelihood will help indicate the probability that a potential program belongs to any of the device models. For each of the obtained variants, the likelihood function is evaluated using a similarity measure. The device model with a maximum likelihood will be the device assigned to the potential program.

The empirical density shape of each water use time and consumption tend to be asymmetric—not surprising since none of these variables can take values lower than zero—and similar to an exponential decaying function. This distribution is very common in practice, and can be modeled as a Log-Normal parametric distribution. This probability family corresponds to situations where the logarithmic transformation of the data behaves as a Normal distribution; and thus the logarithmic transformation was applied to all time and consumption variables. It was also applied to variable X_5 (number of water uses) that reflects a counting process (a usual practice in counting variables; and to variable X_4 (mean flow) to enhance its symmetry and Normal distributed behavior.

Some of the transformed variables have bimodality. This indicates a mixed use pattern in the same household. But as a general approach, a Multivariate Gaussian model (Tong 1989) is assumed for the vector of variables. For simplicity and assurance of a more robust model, correlations among variables are avoided. The model for the program description is marginally defined by its two first centered moments (mean and variance) in each component. The number of model parameters is 20 (two parameters for each component of the vector of descriptors).

General Model By considering a common model for all the households, the Random Vector of Descriptors for use *u*, occurrence *j* has a multivariate Gaussian distribution:

$$(X_1,\dots,X_{10})_{uj} \sim N\left((\mu_{u1},\dots,\mu_{u10}), \Sigma = diag\left(\sigma^2_{u1},\dots,\sigma^2_{u10}\right)\right)$$

For a probabilistic model like this, a measure of the likelihood of one sample occurrence is the Normal density function applied to the values of the vector of statistics. If $_u = (\mu_{u1},\dots,\mu_{u10},\sigma_{u1},\dots,\sigma_{u10})$ represents the set of parameters describing use u for the general level, the likelihood for a specific program can be evaluated by this expression:

$$L\left(\Theta_u;\widetilde{X}_{uj}\right) = f\left(\widetilde{X}_{uj};\Theta_u\right) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi\sigma^2_{ui}}} e^{-\frac{(x_{ui}-\mu_{ui})^2}{2\sigma^2_{ui}}}$$

The parameters were estimated by Maximum Likelihood criterion (Casella and Berger 2001) with the sample programs. For the General Model, all programs of each use are included. On the other hand, to estimate the Specific Level parameters, only the programs of each use for the specific household have been considered. The Maximum Likelihood estimators in this case correspond to the sample mean and variance for each component, calculated over the log-transformed data.

Although the Likelihood function has no units, it is possible to compare several parameters for a random vector in order to establish which of the different parameters it most likely corresponds to. So, for each potential program (artificial combination of water uses), the likelihood for all uses is calculated and the maximum value obtained indicates which use is more probable. The algorithm for assigning use to each program is based on this evaluation of all the artificial programs.

5 Results

The algorithm can be evaluated by comparing the microcompenents it identifies with the devices identified by the loggers.

Table 10 Sensitivity table

Real Water Uses		CORRECT	WRONG	Total
Internal Tap	Water Uses	8721	2611	11332
	% Water Uses	76,96 %	23,04 %	100,00 %
	Consumption	4715,112	6449,516	11164,628
	% Consumption	42,23 %	57,77 %	100,00 %
Cistern	Water Uses	1932	1145	3077
	% Water Uses	62,79 %	37,21 %	100,00 %
	Consumption	10514,491	6598,271	17112,762
	% Consumption	61,44 %	38,56 %	100,00 %
Shower	Water Uses	655	606	1261
	% Water Uses	51,94 %	48,06 %	100,00 %
	Consumption	19507,844	1987,901	21495,745
	% Consumption	90,75 %	9,25 %	100,00 %
Washing Machine	Water Uses	177	398	575
	% Water Uses	30,78 %	69,22 %	100,00 %
	Consumption	968,555	1658,994	2627,549
	% Consumption	36,86 %	63,14 %	100,00 %
Dishwasher	Water Uses	268	268	536
	% Water Uses	50,00 %	50,00 %	100,00 %
	Consumption	504,891	388,697	893,588
	% Consumption	56,50 %	43,50 %	100,00 %
Total of Programs		11753	5028	16781
Total of % Programs		70,04 %	29,96 %	100,00 %
Total of Consumption		36210,892	17083,379	53294,271
Total of % consumption		67,95 %	32,05 %	100,00 %

The efficacy of the assignation process can be summarized in two different tables, where the identified water uses (tables reflect water uses instead of programs, because it is a better unit for making comparisons) are classified into two groups: correctly or wrongly identified. The two tables reflect two different ways of evaluating the results:

Sensitivity Proportion of real water uses that the process has classified correctly.
Specificity Proportion of the water uses that are well-classified.

The following tables provide the sensitivity and specificity results, using the specific model for each household:

Table 10 shows that the process correctly identifies 70 % of the water uses that represent 67.8 % of the monitored consumption. This percentage is the average of all devices. Internal taps and cisterns constitute the maximum rates with 76.9 % and 62.8 %, respectively. In spite of the fact that only 51.9 % of showers are well classified, they consume 90.75 % for this specific device. The minimum rates belong to devices with long programs: 50 % of dishwashers and 30.78 % of washing machines.

Table 11 shows that for all the programs identified as Internal taps, 91.37 % are internal taps. This percentage is very high because the internal taps category works as an absorbent state, including more water uses than would be correct. 72.4 % of all identified cisterns are correct and in the case of showers this percentage decreases to 49.03 %. From this point of

Table 11 Specificity table

Programs identified as…		CORRECT	WRONG	Total
Internal Tap	Water Uses	8721	824	9545
	% Water Uses	91,37 %	8,63 %	100,00 %
	Consumption	4715,112	1056,291	5771,403
	% Consumption	81,70 %	18,30 %	100,00 %
Cistern	Water Uses	1932	736	2668
	% Water Uses	72,41 %	27,59 %	100,00 %
	Consumption	10514,491	3299,119	13813,610
	% Consumption	76,12 %	23,88 %	100,00 %
Shower	Water Uses	655	681	1336
	% Water Uses	49,03 %	50,97 %	100,00 %
	Consumption	19507,844	6377,045	25884,889
	% Consumption	75,36 %	24,64 %	100,00 %
Washing Machine	Water Uses	177	78	255
	% Water Uses	69,41 %	30,59 %	100,00 %
	Consumption	968,555	325,986	1294,541
	% Consumption	74,82 %	25,18 %	100,00 %
Dishwasher	Water Uses	268	2709	2977
	% Water Uses	9,00 %	91,00 %	100,00 %
	Consumption	504,891	6024,937	6529,828
	% Consumption	7,73 %	92,27 %	100,00 %
Total of Programs		11753	11753	5028
Total of % Programs		70,04 %	70,04 %	29,96 %
Total of Consumption		36210,892	36210,892	17083,379
Total of % consumption		67,95 %	67,95 %	32,05 %

view, washing machine identification improves, with 69.41 % of their water uses classified well. Dishwashers are problematic, since only 9.0 % of the water uses associated to dishwashers are correct.

6 Concluding Remarks

The problem of assigning microcomponents to devices based only on the water meter reading is very difficult because of the scarcity of information available: the water consumption time series and the variety of patterns in which people use different devices.

The method proposed is an algorithm that characterizes the water consumption of each device observed in Barcelona as initial parameters and correctly identifies programs and consumption about 70 % of the time, which in these circumstances can be considered high. In our opinion, and given the variability already commented upon, it will be very difficult to significantly increase this percentage without increasing in one way or another information available for making the assignation.

The efficacy of the assignation process is summarized from two different points of view: the sensitivity and the specificity of the process. The sensitivity measures the proportion of devices that the process has classified correctly. The specificity refers to the proportion of the classified devices that are well classified.

The algorithm correctly identifies the devices between 70 % and 80 % of the time, or even higher, depending on how well the chosen parameters reflect the household consumption patterns. Considering the high variability of water consumption patterns—both between households and within each household (different persons)—and the fact that uses are characterized only on the basis of the aggregate consumption provided by the water meter, the results are quite satisfactory.

The algorithm has been implemented in a software program that has been optimized for processing time. The latest version takes between 5 and 10 min to process the consumption time series corresponding to 1 day of one household.

An additional feature of the devised method is that the water identified uses can be truncated according to the value of their likelihood. The likelihood can be interpreted as a measure of the probability that the assigned use is correct, so a minimum level of certainty can be defined. This means not classifying a certain percentage of water uses, let's say the 10 % of those who have a lower likelihood (higher "probability" of being incorrect) in exchange for greater certainty of correct classification in the remaining 90 %.

Once the Barcelona project was finished, the methodology was applied to ten houses in Murcia. The results were satisfactory obtaining a 71.8 % of the programs well identified, representing the 60.8 % of the total water consumption. The houses were selected using variability information. The experience from working in Barcelona informed the various changes introduced into the algorithm for improving the sensitivity and specificity of identification. The primary aims were to increase the general percentage of identification, improve the results for the devices with non-continuous water use, and to introduce a stage of learning procedure into the algorithm. The Murcia application was a validation stage that established a general methodology to be used in other municipalities.

In what follows, we briefly comment on possible ways to further improve the algorithm, which will hopefully result in small improvements in assignation processing and also in ways to increase the information available for making the assignment.

Possible algorithm improvements:

- Further investigation of variables that will help characterize long programs (washers and dishwashers) and thus, hopefully, have a higher correct identification rate.
- Devise ways to incorporate information regarding the number of persons living in the household
- Although, in applying a general model to all households, we found no significant improvements in the assignment process by using device models specific to each household, a possible (however complicated and uncertain) improvement may be to introduce some type of learning procedure in the algorithm, so that it adapts itself to the household habits. This would be very difficult without human intervention.
- The group of internal taps presents very high variability; so it is not surprising that it absorbs water uses that belong to other devices, especially showers, dishwashers and washing machines. At this point, it is not clear how to mitigate this fact. It may not even be possible, but it certainly would be a nice improvement.

Acknowledgements The authors are grateful to R+I Alliance for the financial support that made it possible to develop this project. The authors are also grateful to AQUAGEST SOLUTONS for their very useful comments and suggestions during the preparation of this manuscript.

References

Casella G, Berger LR (2001) Principle of data reduction. In: Thomson Learning (ed) Statistical inference, 2nd edn., United States, pp 271–310

Eurostat (2007) Consumers in Europe – Facts and figures on services of general interest http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-DY-07-001/EN/KS-DY-07-001-EN.PDF. Accessed 16 April 2012

Farinaccio L, Zmeureanu R (1999) Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. Energy and Buildings 30:245–259

Fontdecaba S, Grima P, Marco L, Rodero L, Sánchez-Espigares J, Solé I, Tort-Martorell X, Demessence D, Martínez De Pablo V, Zubelzu J (2012) A methodology to model water demand based on the identification of homogenous client segments. Application to the city of Barcelona. Water Resources Management 26:499–516. doi:10.1007/s11269-011-9928-5

Mayer WP, William B, DeOreo TE, Lewis DM (2003) Residential indoor water conservation study: evaluation of high efficiency indoor plumbing fixture retrofits in single-family homes in the East BayMunicipal Utility District Service Area. Prepared to East BayMunicipal Utility District and The United States Environmental Protection Agency

Richter C, Stamminger R (2012) Water consumption in the kitchen –a case study in four European countries. Water Resources Management 26:1639–1649. doi:10.1007/s11269-012-9976-5

Tong YL (1989) The multivariate normal distribution. Springer Series in Statistics, 1st edn., United States

Yamagami S, Nakamura H (1996) Non-intrusive submetering of residential gas appliances. ACEEE Summer Study on Energy Efficiency in Buildings, 265-273