Correlating water consumption behaviour with billing, infrastructure and socio-demographic factors

D. Loureiro, S. T. Coelho, M. Rebelo, P. Ramalho & H. Alegre
National Civil Engineering Laboratory, Av. Brasil, 101, 1700-066, Lisbon (Portugal)

D. I. C. Covas
Instituto Superior Técnico - Technical University of Lisbon, CEHIDRO - Department of Civil Engineering, Av. Rovisco Pais, 1049-001, Lisbon (Portugal)

A. Pina
SMAS Oeiras-Amadora Water Utility, Oeiras (Portugal)

ABSTRACT: This paper describes a study of residential consumption characterization at the district metering area (DMA) level, taking into consideration billing, socio-demographic and infrastructure-related factors. The article describes the methodology used and presents the results of its application to a set of 20 network districts in the Lisbon (Portugal) metropolitan region, over a period of 12 months of continuous monitoring, including the analysis of consumption scenarios; the estimation of daily consumption patterns; and the study of the most relevant factors that influence consumption. The modelling of consumption patterns involved the analysis of large consumers, the identification of consumption scenarios and the estimation of daily consumption patterns. Multivariate statistical analyses were carried out to obtain factors and identify the correlations with consumption statistics. The main correlations with daily, weekly and seasonal statistics are shown.

1 INTRODUCTION

Detailed knowledge of water consumption is essential for the best management and operation of a water utility, both in terms of common statistics (e.g., minimum night flows, instantaneous, daily and monthly peak consumption), and as regards long-term, seasonal, weekly and daily behaviours. Profiling minimum night flow is crucial in controlling water losses (Almandoz *et al*. 2005; Buchberger and Nadimpalli 2004); peak consumption values and their timing are important for network design or household flowmeter selection (Javier *et al*. 2008; Orazio *et al*. 2009); demand prediction allows for better scheduling of pumping to service reservoirs, taking advantage of lower cost power tariffs; water quality models are strongly dependant on the water demand paths incorporated in the hydraulic model (Blokker *et al*. 2009); detailed daily behaviour yields some of the best available insights into urban social patterns on a continuously monitorable basis (Corral-Verdugo *et al*. 2003; Domene and Saurí 2006); and a solid knowledge of consumption is essential in implementing measures to improve efficient water use (Dias *et al*. 2007).

The assessment of the most important socio-demographic and technical factors influencing the various components of human consumption is essential in understanding and predicting its behavior – and is particularly useful when extrapolating for networks and expansions where flow data are not available, either because they are unmetered or simply because they haven't been built yet (Loureiro *et al*. 2006; Warren and Cunningham 2007).

Different approaches have been deployed for water consumption modelling, by varying the spatial scale (i.e., size of network or population served) and the time discretisation (e.g., 1 second, 1 minute, 15 minutes), depending on the purpose. Buchberger and Wells (1996) concluded that the behaviour of demand over time can be represented by a combination of rectangular pulses, which is the basis of many models that estimate daily demand pattern and night use at household level (Blokker *et al*. 2009; Garcia *et al*. 2003; García *et al*. 2004). At network district metering area (DMA) level, different statistical distributions have been used to simulate each time of day (Coelho 1988; De Marinis *et al*. 2006). Modelling of peak water demand and characterization of factors that can influence it is fundamental for system design (Tricarico *et al*. 2005; Zhang *et al*. 2005).

This paper presents the results for the first 12 months of a consumption study in the cities of Oeiras and Amadora, in the Lisbon (Portugal) metropolitan area, with the purpose of establishing the most important factors influencing consumption at the network level, for the systems involved. A set of 20 residential urban network district areas was selected, ranging from 1000 to 5000 properties served, with a variety of consumption behaviours and socio-demographic contexts. The network districts are all continuously monitored and the study benefited from the availability of fully updated, calibrated network models of all areas concerned.

The main scientific contribution of the research presented in this paper is the systematic approach developed for water demand characterisation, particularly at the DMA level.

2 METHODOLOGY

A general methodology has been developed (Loureiro 2006) to analyse consumption in water distribution systems at different levels of spatial (e.g., DMA, individual property) and time discretisation (e.g., 1 min, 15 min). This methodology, presented in Figure 1, is divided into 5 main modules. Each module can be further detailed depending on the level and objective of consumption analysis (e.g., Leakage assessment, demand estimation, pressure management).

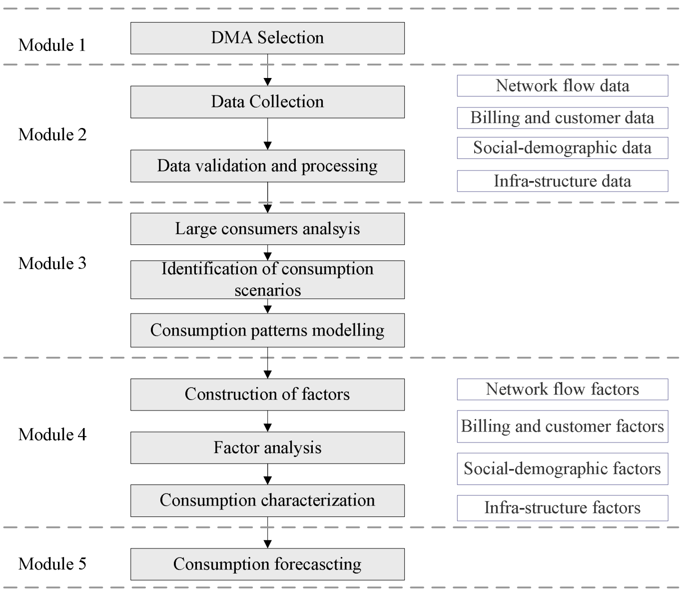


Figure 1 – General methodology.

In this paper, the methodology is applied at DMA level. This is part of a broader research effort that includes the application of this methodology to a larger number of case studies nationwide, covering areas with widely varying urban, infrastructural and socio-demographic characteristics. This effort will start a national database of demand patterns and data as a function of socio-demographic and infrastructural factors.

Module 1 of the methodology establishes the requirements for DMA selection in terms of users (number and type), metering (i.e., meter accuracy, type of users metered), network (i.e. operating conditions), and consumption data (i.e., time step and duration).

Module 2 is where data collecting, validating and processing takes place. Additionally to network flow data, water utilities provide GIS-referenced network data (no. household connections, pipeline type and condition, pressure levels, pipe and household connection failure rates, among others) and annual billing and customer data. Socio-demographic data, related to buildings, households and families, is obtained from the current census (2001), publicly available as a GIS data layer. A GIS tool has been developed to intersect the census's statistical units with the area served by each DMA. Network flow data is validated using automated algorithms to filter out abnormal flows (such as negative values, parameterization problems, pipe breaks) and flow records are standardised in 15 minutes intervals (i.e., 96 values per day).

In Module 3, consumption is modelled in order to obtain daily consumption patterns and flow statistics such as peak flow values, minimum night flow and average flow. After processing all time series in Module 2, water balances are performed in the case of DMA with multiple sources or in the cases where large consumers exist and their detailed flow records are available. Pattern analyses, using cluster analysis, were carried out for the identification of the main consumption scenarios, mainly in terms of seasonal, weekly and daily patterns.

In Module 4, a set of infrastructural, billing and be customer, and Census data variables has been selected. Principal Component Analysis (PCA) was used in order to reduce the number of variables to be related with consumption, and to reduce or eliminate information redundancy. Relations between factors (i.e., new variables) and consumption statistics have been analysed using correlation matrix, PCA and cluster analysis.

Module 5 aims to forecast daily consumption patterns based on the most relevant factors. This paper presents results obtained in Modules 1 through 4.

3 CASE STUDY SELECTION

As previously mentioned, Module 1 imposes a set of requirements for DMA selection. The DMA must be predominantly residential for each consumption scenario (household billed consumption > 70% total consumption) and must correspond to stabilized urban areas. Any large consumer (i.e., those whose consumption is likely to have a significant effect on total inflow and on its daily pattern) must be individually metered.

Total inflow must be metered through reliable and adequately sized and fitted flowmeters. The boundaries of the network district must be clearly defined and watertight (either permanently or during data collection). Operating configurations must be registered and kept constant in all metering periods (such as, e.g., for different seasons) to ensure data consistency.

The recommended flow monitoring frequency is 10 to 15 minutes, recording either total or average flow over the time step. That frequency is thought to be sufficiently detailed for a correct depiction of daily variation and identification of a good set of behaviour features. Shorter time steps will yield higher detail but also dramatically increase data storage requirements. One limitation of using a 10 to 15 min time step is the loss of detail and the difficulty in estimating real losses from night flows, as the time step is too large to assess the actual leakage background flow.

The study involved the analysis of twenty DMA, in the cities of Oeiras and Amadora, located in the greater Lisbon metropolitan area (Portugal), ranging between 1000 and 5000 households served. The metered network districts are mostly made up of plastic pipes (Figure 2), with recorded pipe failure rates of between 14 and 88/100 km/year and household connection failure rates of between 2 and 57/1000 household connections/year. Household consumption varies between 222 and 420 l/household/day.

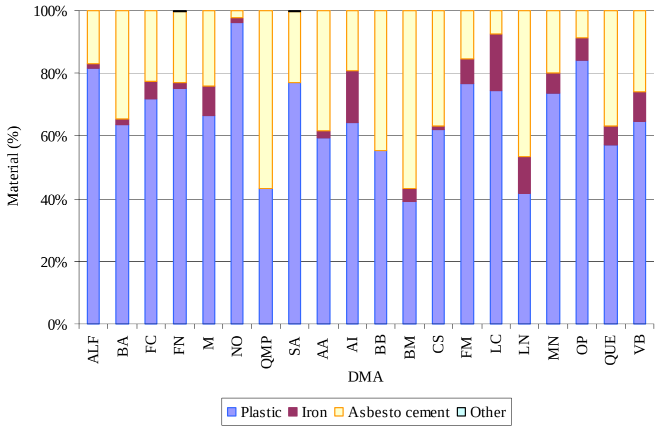


Figure 2. Pipe material in DMA.

The study involved the collection of information which is not usually analyzed together, with important differentiation in time and spatial aggregation. Twelve months of network flow records (with 10-15 min time steps) were used, together with detailed billing data obtained from bi-monthly readings (household and non-household consumption) for the same period. Socio-demographic data for the populations supplied were obtained from the 2001 Census database, available on a GIS georeferenced format as previously mentioned.

GIS pipe network and service connection data, and information about pipe and service connections failures were also collected. The variety of sources of information underpinned the need for enforcing a rigorous set of data requirements, fundamental for methodology validation.

4 MODELLING OF CONSUMPTION PATTERNS

4.1 *The impact of large consumers*

Large consumers, such as schools, hotels or a local market, can have a significant impact on the total billed consumption and exhibit a daily pattern that is completely different from residential users. Identification and metering of large consumers is fundamental because they can influence the overall DMA consumption pattern and get in the way of the accurate characterization of patterns or the use of algorithms for automated filtering out of abnormal peaks (such as those due to pipe breaks).

A set of 22 large consumers with 12 months of consumption data was analysed. Large consumers were grouped in six categories (Figure 3). Large consumers represent 2 to 10% of the total DMA billed consumption; their individual consumption ranges between 500 and 3500 l/client/day, showing a wide variation even inside each category.

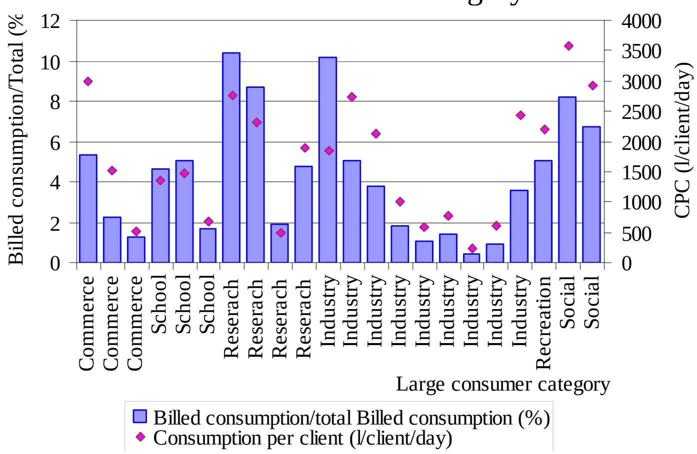


Figure 3. Billed consumption and consumption per client for large consumers.

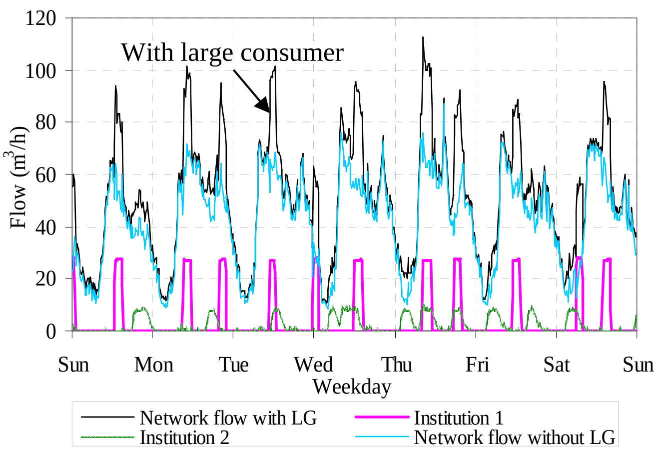


Figure 4. Network flow and large consumers daily time series for DMA “M”.

Figure 4 shows the impact of large consumers on the network flow in a sample district (DMA “M”), where the large consumers represent 15% of total annual billed consumption; this behaviour is completely different from that observed in DMA “NO”

(Figure 5), where large consumers represent 19% of total annual billed consumption, but their consumption is more homogeneously distributed along the 24 hours. Consumption scenarios for each large consumer were identified using cluster analysis and daily consumption patterns where estimated. The daily patterns of large consumers are essential for network modelling.

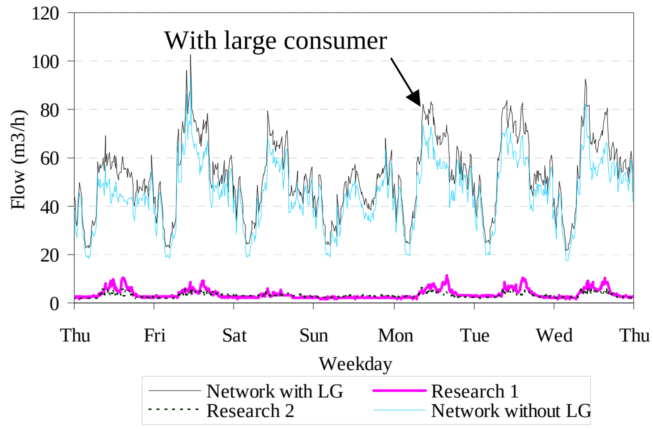


Figure 5. Network flow and large consumers daily time series for DMA "NO".

4.2 *Daily consumption patterns*

One of the goals of pattern analysis is the identification of homogeneous and representative periods to define pattern scenarios. In a short term (i.e., annual base) patterns can vary throughout the year, influenced by seasonal effects, the week or the day. Cluster analysis was used to group months, week and instantaneous behaviour using hourly data. Agglomerative hierarchical clustering has been adopted and different distance measures and types of linkage were tested.

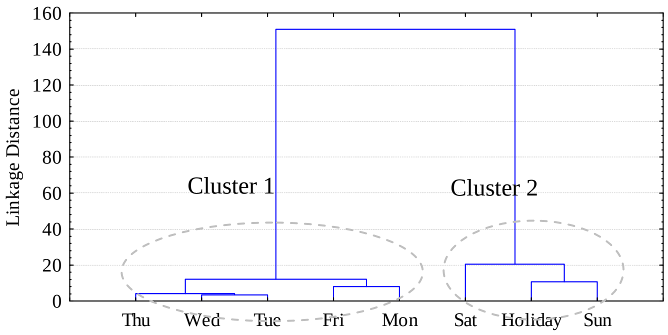


Figure 6. Dendrogram for categories of "day" variable for DMA "CS" during January, February and December.

Figure 6 shows the results of agglomerative hierarchical clustering using Euclidean distances and Ward's method as type of linkage. The dendrogram clearly shows two clusters for the "day" variable: one for weekdays and another for weekend and holidays. Daily consumption patterns are shown in Figure 7.

Daily consumption patterns are a prime tool for studying and translating user behaviour and water use variations throughout the day. Network modelling and several other engineering and planning analyses, such as used for design or operational purposes, depend heavily on this key element and associated flow statistics, such as peak factors. The determination of such daily consumption patterns is based on the statistical processing of continuous network flow records, in order to obtain the average curve over the 24 hours of the day.

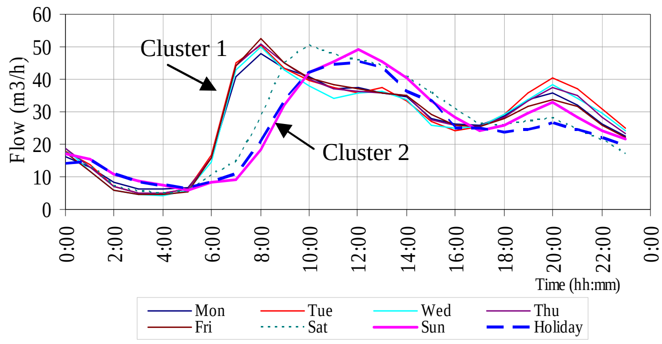


Figure 7. Daily consumption patterns by weekday for DMA "CS" during January, February and December.

Each point of a daily curve represents all the consumption values recorded at that particular time of the day, for that particular day of the week. If long-term trends are removed, consumption behaves as a simple random variable. Statistics of random variables may then be applied, in order to test if characteristics and trends identified in the observation stage are meaningful. Coelho (1988), using consumption data from urban areas near Lisbon, has shown that these cross distributions tended to present several peaks and be rather scattered; this effect seems to attenuate significantly when the raw flow data is divided by its corresponding daily average, producing a dimensionless pattern (Figure 8).

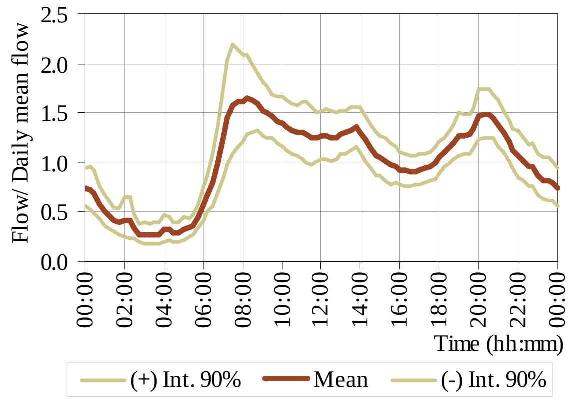


Figure 8. Dimensionless daily consumption pattern for DMA "AA" during weekdays.

To complement the mean curve, confidence bands, to express dispersion at each point have been obtained. Following Coelho (1988), the log-normal distribution was retained for the computation of daily patterns in this study.

Daily, weekly and seasonal statistics flows have been calculated. Figure 9 shows the variation of peak factors (monthly, daily and instantaneous) for the set of twenty DMA analysed. Monthly and daily peak factors are homogeneous across the analysed DMA, while instantaneous peak factors have greater variability. Comparing with design values, the difference is not very significant in terms of instantaneous peak factors, with exception of three DMA: ALF, CS and LC.

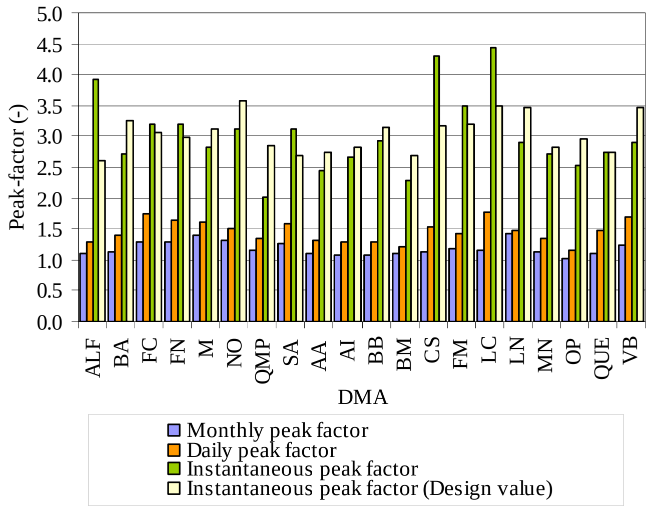


Figure 9 – Peak factors in DMA.

5 FACTOR ANALYSIS

Previously to consumption characterization a set of variables were obtained from infrastructural, socio-demographic, billing and customer information at the DMA level. Due to the large number of variables obtained, it is important to aggregate information into new more synthetic dimensions and at the same time detecting previously redundant variables without losing significant information. To achieve this goal, Principal Component Analysis (PCA) has been performed. The number of factors to be retained was established based on two criteria: Eigen values >1 and total variance > 80%. In order to obtain a clear pattern of variables that contribute to the new factors, that is, factors that are somehow clearly marked by high loadings for some variables and low loadings for others the Varimax Rotation Method was used.

Figure 10 shows the projection of infrastructure variables on the Factors 1-2. These new factors allow an explanation of 80% of initial variance. According to Factor 1, pipe age and plastic pipes rate are positively associated and both vary inversely with asbestos cement pipe rate. Thus, this factor, that compares DMA with younger pipes and with a higher proportion of plastic pipes to DMA with a higher rate of asbestos cement pipes was designated the “pipe material factor”. Factor 2 - service connection - associates network with large diameter to a high density of irrigation service connections and opposites these variables to a high density of service connections and a high rate of iron pipe. These new variables were used to correlate with consumption variables.

Pipe and service connection failures were not used for the construction of these factors, since most of water utilities do not have historical data desegregated by DMA and categorized by pipe and per service connection.

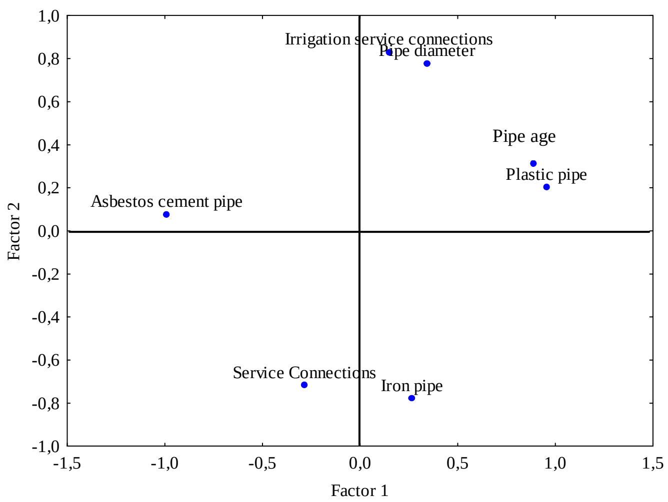


Figure 10. Projection of infrastructure variables on Factors 1 and 2 (Varimax Rotation Method).

Figure 11 presents the projection of cases on factors 1-2 and allows a characterization of the DMA regarding infrastructure. Group I identifies DMA with a high rate of asbestos cement pipes, whereas Group II matches DMA with a low density of service connections and rate of iron pipes, and a high rate of irrigation service connections and higher pipe diameter. Group III corresponds to DMA with recent pipes and high rate of plastic pipes. Group IV corresponds to DMA with a high density of service connections and high rate of iron pipes.

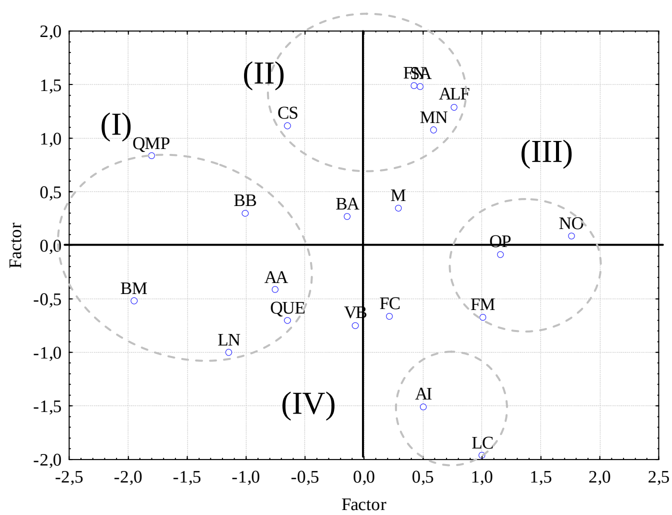


Figure 11. Projection of DMA on Factors 1-2 (varimax rotation method).

Although, pipe and service connection failures were not used for Factors construction, it was verified that DMA with a high rate of asbestos cement pipes have a high rate of service connection failures

(Figure 12). Regarding pipe failure it was not possible to find any reliable correlation with other infrastructure variables.

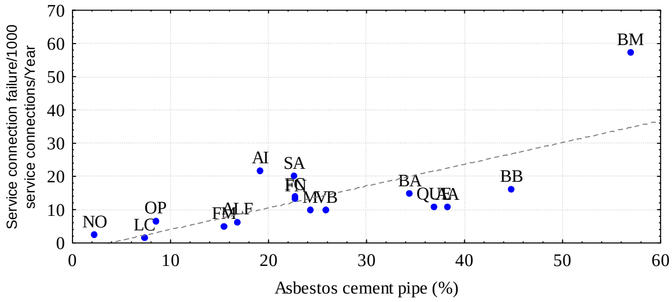


Figure 12. Characterization of DMA considering asbestos cement pipe rate and service connection failure.

Table 1 summarizes the factors obtained for infrastructure and socio-demographic categories and the variables that contribute for each factor.

Table 1. Factors composition per category and by input variables.

Category	Factors	Input variables
Infrastructure	Pipe Material	Pipe age (year) Plastic pipes (%) Asbestos cement pipes (%)
	Service connection	Irrigation service connections (%) Pipe diameter (mm) Service connections density Iron pipes (%)
Socio-demographic	Building Age	Buildings with age >27 years (%) Buildings with age < 16 years (%)
	Building height	Buildings with 1-2 floor (%) Buildings with > 5 floor (%)
	Social stratification	University graduates (%) Economic mobility (%), Population with ≤ 9 years of education(%), Population working in the municipality (%)
	Age	Population above age 65 years(%) Inactive workers (%)
	Family structure	Families with adolescents (%) Families with elderly (%) Small families (1- 2 elements) (%) Medium Families (3-4 elem.) (%)

6 WATER CONSUMPTION CORRELATIONS

Statistical multivariate analyses, such as correlation matrix, PCA and cluster analysis, were used to identify relations between consumption and the factors presented previously. Figure 13 illustrates the variation of monthly peak factor against the social stratification factor.

This factor increases with the rate of university graduates and economic mobility, and decreases with population with > 9 years of education and working locally in the municipality. The correlation is positive, which indicates that DMA with high peak factors are associated with a population with a high rate of university graduates and high mobility.

Figure 14 shows the variation of minimum consumption with material factor. Material factor increases with more recent pipe networks and the rate plastic pipes and decreases with the rate of asbestos cement (AC) pipes. The correlation is negative, indicating that high values of minimum night flow area are associated to DMA to a high rate of AC pipes. Figure 15 shows daily consumption patterns for DMA "ALF" and DMA "FN" for week days.

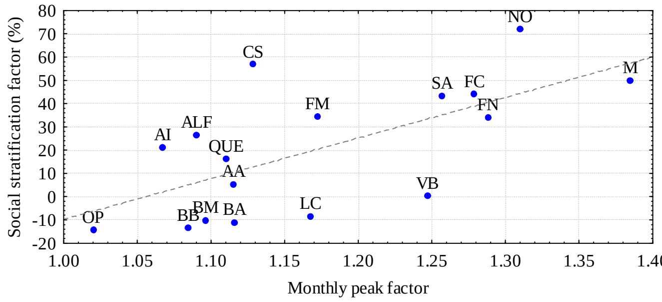


Figure 13. Characterization of DMA considering Monthly peak factor and social stratification factor.

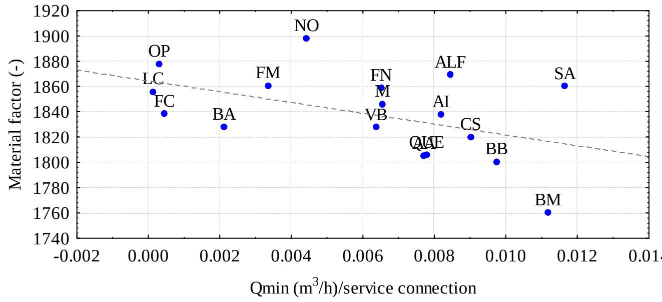
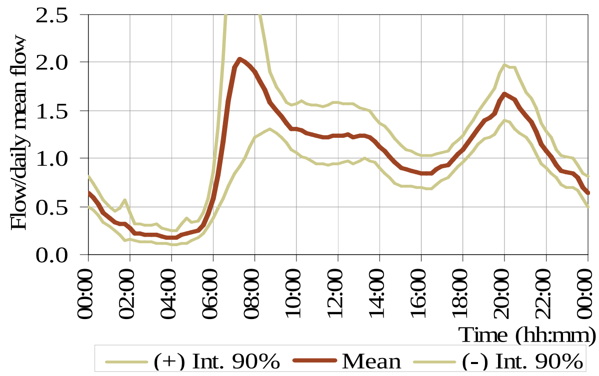
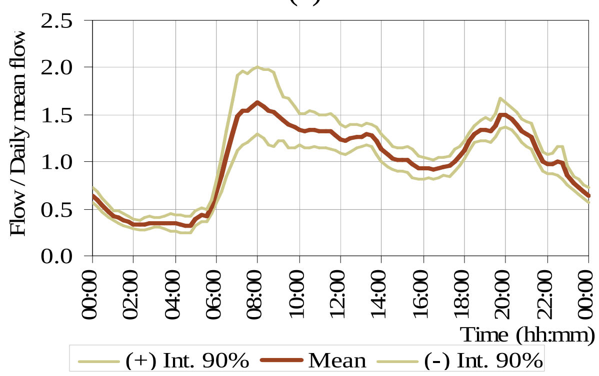


Figure 14. Characterization of DMA considering minimum night flow and the pipe material factor.



(a)



(b)

Figure 15. Daily consumption pattern for (a) DMA "FN" and (b) DMA "ALF" during weekdays.

Figure 15 illustrates how consumption patterns can exhibit significant differences in terms of instantaneous peak factors, consumption during lunch period and confidence bands. DMA “ALF”, characterized by a low age factor (rate of population above 65 years and inactive workers), shows high instantaneous peak consumption values during the morning and at dinner time, low consumption values during the lunch period and higher dispersion along the day. DMA “FN”, characterized by a high rate of small families (2 or less people), described by the family structure factor, and older population than DMA “ALF”, shows a more homogeneous behaviour during the day with lower instantaneous peak consumption values, higher consumption during the lunch period and lower dispersion along the day.

7 CONCLUSIONS

The case study of Oeiras and Amadora showed that the methodology developed is robust and allows for accomplishing the objectives of demand characterisation:

- analysis of consumption scenarios;
- estimation of daily consumption patterns;
- study of the most relevant factors that influence consumption.

The main difficulties encountered, which may be relevant and should be taken into account in other similar applications, include:

- large consumers (i.e., with consumption significantly above average domestic consumers) have a major impact on the DMA consumption pattern and must be dealt with individually;
- national surveys have a low frequency, and therefore census socio-demographic data may be outdated, particularly in fast growing areas; this may require the use of other sources of information.

The application of this methodology to other case studies is necessary and is currently under way, in order to consolidate the results and create a robust and rich database that will allows for extrapolation to areas where consumption data are not available or reliable.

8 ACKNOWLEDGEMENTS

The authors wish to acknowledge SMAS Oeiras & Amadora for providing the data and the insights essential to the analysis. Dália Loureiro is grateful to the Portuguese Foundation for Science and Technology (FCT) for the PhD grant (SFRH/BD/23819/ 2005). Pedro Ramalho’s research scholarship was equally funded by FCT (project grant no. POCI/ECM/60917/2004).

REFERENCES

Almandoz, J., Cabrera, E., Arregui, F., Cabrera, J. E., and Cobacho, R. (2005). "Leakage Assessment through Water Distribution Network Simulation." Journal of Water Resources Planning and Management, 131(6), 458-466.

Blokker, E. J. M., Vreeburg, J. H. G., and Dijk, J. C. v. (2009). "Simulating Residential Water Demand with a Stochastic End-Use Model." Journal of Water Resources Planning and Management.

Buchberger, S. G., and Nadimpalli, G. (2004). "Leak Estimation in Water Distribution Systems by Statistical Analysis of Flow Readings." Journal of Water Resources Planning and Management, 130(4), 321-329.

Coelho, S. T. (1988). "A System for Demand Analysis and Forecasting in Water Supply Systems," Master Thesis, University of Newcastle.

Corral-Verdugo, V., Bechtel, R. B., and Fraijo-Sing, B. (2003). "Environmental beliefs and water conservation: An empirical study." Journal of Environmental Psychology, 23(3), 247-257.

De Marinis, G., Gargano, R., and Tricarico, C. (2006). "Water demand models for a small number of users." 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati.

Dias, S. L., Lucas, H., Almeida, M. C., Alegre, H., and Viriato, M. (2007). "Identification of opportunities to improve efficiency by water consumption assessment." Water Science and Technology: Water Supply, 7(4), 57-63.

Domene, E., and Saurí, D. (2006). "Urbanization and water consumption: influencing factors in the metropolitan region of Barcelona." Urban Studies, 43(9), 1605-1623.

Garcia, V. J., Cabrera, E., Garcia-Serra, J., Arregui, F., and Almandoz, J. (2003). "Stochastic prediction of the minimum night flow demand in a district metered area." Advances in Water Supply Management, A. A. Balkema, Imperial College, Londres, Reino Unido, 665-672.

García, V. J., García-Bartual, R., Cabrera, E., Arregui, F., and García-Serra, J. (2004). "Stocastic model to evaluate residential water demands." Journal of Water Resources Planning and Management, 130(5), 386 - 394.

Javier, M.-S., Pedro, L. I.-R., Rafael, P.-G., and Lopez-Jimenez, P. A. (2008). "Hydraulic Analysis of Peak Demand in Looped Water Distribution Networks." 134(6), 504-510.

Loureiro, D. (2006). “Consumption analysis methodologies for the efficient management of water distribution systems”, PhD Thesis program, Instituto Superior Técnico, Lisbon Portugal (in Portuguese).

Loureiro, D., Coelho, S. T., Machado, P., Santos, A., Alegre, H., and Covas, D. (Year). "Profiling residential water consumption." Water Distribution Systems Analysis Symposium, University of Cincinnati, Cincinnati, Ohio, USA.

Orazio, G., Daniele, L., and Andrew, F. C. (2009). "Deterministic versus Stochastic Design of Water Distribution Networks." ASCE, 117-127.

Tricarico, C., de Marinis, G., Gargano, R., and Leopardi, A. (2005). "Peak water demand for small towns." 8th International Conference on Computing and Control for the Water Industry " Water Management for the 21st Century", Center for Water Systems, University of Exeter, Exeter, UK, 113-118.

Warren, R., and Cunningham, A. (2007). "Leakage Methodology Review: variation in per capita consumption estimates." J0676\GD04\03, Tynemarch Systems Engineering Ltd.

Zhang, X., Buchberger, S. G., and van Zyl, J. E. (Year). "A Theoretical Explanation for Peaking Factors." World Water Congress 2005, ASCE, Alaska, USA, 51-51.