Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 70 (2014) 1063 – 1073

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

Spatial and temporal forecasting of water consumption at the DMA level using extensive measurements

A. Mamade[a*], D. Loureiro[a], D. Covas[b], S.T. Coelho[a], C. Amado[c]

[a]*National Civil Engineering Laboratory, Lisbon, Portugal*
[b]*CEHIDRO, Instituto Superior Técnico (IST) – Technical University of Lisbon, Lisbon, Portugal*
[c]*CEMAT, Instituto Superior Técnico (IST) – Technical University of Lisbon, Lisbon, Portugal*

Abstract

Reliable water forecasting provides the basis for making operational, tactical and strategic decisions for water utilities. The aim of this paper is to present the results of a spatial and temporal forecasting of water consumption patterns that has been carried out at the District Metered Area (DMA) level. Flow, infrastructure and billing data were collected from water utilities and combined with recently published census data. Results include empiric relations for estimating design and operational parameters for different seasonal and weekday scenarios, as well as average demand patterns for different consumer profiles taking into consideration key variables.

Selection and peer-review under responsibility of the CCWI2013 Committee

Keywords: District Metered Area (DMA), demand forecasting, regression analysis, spatial analysis, water consumption

1. Introduction

Measuring water consumption at the District Metering Area (DMA) level is becoming a current practice in most water utilities with the recent development of metering equipment and Supervisory Control And Data Acquisition (SCADA) systems. The characteristics of consumption data can vary considerably between DMA throughout different seasons, weeks and days, due to the number and heterogeneity of consumers and households, the existence of large consumers and the condition and operation of pipe networks.

* Corresponding author. Tel.:+351 913 599 960.
E-mail address: aisha@lnec.pt

1877-7058
Selection and peer-review under responsibility of the CCWI2013 Committee
doi:10.1016/j.proeng.2014.02.118

Forecasting water consumption is challenging due to the nature and quality of available data, the numerous drivers that lay behind water consumption and the multiplicity of forecast horizons and spatial scales involved (Donkor et al., 2012). The influences of economic, socio-demographic and geographic variables as key drivers of water consumption (Grafton et al., 2011; March Corbella and Saurí i Pujol, 2009; Shandas and Hossein Parandvash, 2010) need to be identified through an inter-disciplinary work between engineers, economists, sociologists and urban planners. However, there are still few studies considering this integrated approach. On the temporal forecasting side, few studies consider consumption scenarios on a seasonal or daily basis, despite the reckoned differences between winter and summer behaviours (Polebitski and Palmer, 2010). On the spatial forecasting side, little research has examined the influence of urban land use patterns on water consumption (Shandas and Hossein Parandvash, 2010).

The present paper focuses on developing a spatial and temporal forecasting of water consumption at the DMA level. It uses extensive measurements, since data from DMA distributed throughout the country of Portugal were collected, while previous studies used DMA from a single region (Loureiro, 2010). These extensive measurements from different regions with different consumption profiles allow a spatial forecasting. Furthermore, the collection of flow data with large historical data (i.e., a year) allows identifying different seasonal and weekday scenarios that enable a temporal forecasting. In this study the term "consumption" refers to all the components of urban consumption (domestic, non-domestic, leakage), while the term "flow" data refers solely to flow measurements. The methodology that establishes the approach for consumption forecasting as well as case-studies are presented. A series of empiric relations that predict water consumption in different scenarios is explored.

2. Methodology

The methodology used for spatial and temporal demand forecasting was adapted from Loureiro (2010) and involves a four-step procedure: *Data collection, Data processing, Data characterization and Spatial and temporal consumption forecasting*. As highlighted in **Errore. L'origine riferimento non è stata trovata.**, the present paper focuses in steps 1, 3 and 4.

Prior to data collection, the DMA must fulfill a set of requirements to be valid for spatial and temporal consumption forecasting (Loureiro, 2010; Mamade, 2013).

The first step (*Data collection*) includes collecting data from different DMA, typically available in a water utility, in terms of infrastructure, billed consumption and customers and flow data.

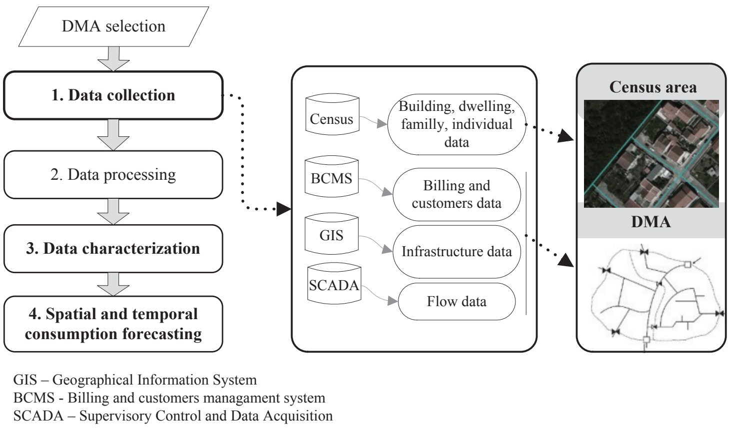

Fig. 1. Methodology for water consumption profiling, adapted from (Loureiro, 2010).

Infrastructure data were provided in GIS system format. Monthly billed consumption data (domestic and non-domestic) were collected with one year time interval. Flow data readings from the utilities' SCADA systems were collected for DMA with continuous flow monitoring. Socio-demographic data concerning building, dwelling, family and individual categories data were downloaded from the last census held in Portugal. These data were downloaded at the smallest census area, that corresponds to homogeneous building and living zones, with approximately 300 households (INE, 2012). A DMA may include several census areas.

Collected data were processed in step 2 *(Data processing*). For flow data, this step involved data validation, normalization and outlier detection and cleaning. In terms of infrastructure, billing and customer data, the databases were standardized and combined in order to have only the necessary information required for the analysis. For socio-demographic data, a geoprocessing tool that calculates the weight of each census area and transforms the statistical descriptors from the census area level into the DMA level was used. This tool was developed by Loureiro (2010) and improved in this study.

The third step (*Data characterization*) involved calculating socio-demographic, infrastructure, billing and consumption variables (Table 1). Socio-demographic variables were calculated using the geoprocessing tool. Variables concerning the building, dwelling, family and individual categories were calculated. A large number of socio-demographic variables are presented herein (22 variables). Most were initially used to characterize the social dimension in other urban infrastructures (e.g., airports) (Rebelo and Freitas, 2004). Loureiro (2010) has shown that these variables were appropriate for correlation with consumption variables in water distribution systems (WDS). Infrastructure variables describe the main characteristics of the pipe network (i.e., material, diameter and installation year) and service connections (i.e., number of service connections and service connection pipe length). Billing variables characterize the domestic and major categories of non-domestic consumption (i.e., commerce-industry, collective and public). In terms of domestic consumption, total domestic consumption within each tariff was also calculated. The tariff categories considered were: 1st tariff – up to 5 m^3/month, 2nd tariff – between 5-15 m^3/month, 3rd tariff – between 15-25 m^3/month and 4th tariff – above 25 m^3/month (IRAR, 2009).

Table 1. Socio-demographic, infrastructure and billing variables.

Socio-demographic				Infrastructure		Billing	
Building	Dwelling	Family	Individual	Service connection	Pipe	Domestic	Non-domestic
Buildings until 1970, 1980, 1990, 2000 and up to 2011 Buildings with 1-2 , 3-4 and ≥ 5 floor	Residential immobility Rented dwellings Vacant dwellings	Families with adolescents Families with elderly Families with unemployed Small families (1-2 elements) Medium families (3-4 elements) Large families (≥5 elements)	Population above age 65 years Inactive workers University graduates Economic mobility Active population mobility Population with 12 years of education	Service connection density (sc/km) Average service connection pipe length (m)	Average installation year (year) Average diameter (mm) Total stainless steel pipe length (%) Total grey iron pipe length (%) Total asbestos cement pipe length (%) Unknown material pipe length (%)	Average domestic consumption (l/inh·day) Total domestic consumption (%) Total domestic consumption within each tariff category (%)	Commerce-industry billed consumption (%) Collective billed consumption (%) Public billed consumption (%)

This step also involved the calculation of consumption variables and patterns for all the DMA taking into consideration the different seasonal and weekday consumption scenarios (Table 2). Seasonal scenarios may be related with changes in outdoor uses throughout the year (e.g., garden watering and swimming pool filling); weekday scenarios may related to water use changes between working days and weekends.

Consumption scenarios were obtained using hierarchical cluster analyses (Ward's method). First, this statistical technique was used to identify groups of months with a similar behaviour. Each month corresponded to a case for the analysis and was characterized by the average consumption pattern. Second, cluster analysis was used again to identify group of weekdays with similar behaviour, within each seasonal scenario and the same approach was adopted. Then, a set of consumption variables was calculated. The focus was to calculate important variables for network operation and water losses control (i.e., minimum night consumption, and average consumption during minimum night consumption period) and for pipe network design and rehabilitation (i.e., peaking factors and average consumption). Therefore, consumption variables based on instantaneous flow time series were calculated for all scenarios (i.e., global, seasonal and weekday). The minimum consumption during the night period (0:00-6:00) was firstly identified and removed from the consumption time series, for a separate analysis and for all consumption scenarios, since it is an important variable to evaluate leakage at the DMA level (Farley and Trow, 2003).

For the subsequent analyses, 49 variables concerning socio-demographic, infrastructure and billing categories and 55 variables related to domestic billed consumption, total consumption and non-revenue water were obtained, along with 6 daily consumption patterns (one for each identified scenario). Thus, this comprehensive approach allows forecasting the different components of urban consumption taking into consideration a broad number of key variables.

The fourth step (*Spatial and temporal consumption forecasting*) involved data reduction using Principal Components Analysis (PCA) and correlation and Multiple Linear Regression (MLR) analyses. PCA was applied to reduce the number of variables from each category – socio-demographic, infrastructure and billing – into Principal Components (PCs). In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO-test) was used. The sample was adequate if the value of KMO was greater than 0.6 and the total explained variance for each category was greater than 75%. After the PCA, simple correlation analysis was carried out to assess possible correlations between PCs and consumption variables. Once identified the most important correlations MLR was carried out (Donkor et al., 2012) by setting consumption variables as dependent variables and predictor variables as independent variables (these include the PCs calculated). In order to ascertain good regressions, the adjusted correlation coefficient (r_a^2) that reflects the number of independent variables as well as the sample size is calculated, as well as the p-value.

In what concerns the average daily consumption patterns, a new cluster analysis was carried out in order to group DMAs with similar patterns. Then, the average patterns for each group were converted into dummy variables and analysed with a correlation matrix using socio-demographic variables. The consumers' behaviour is analysed for different weekday scenarios. Only socio-demographic variables were used, since these were found to be the most important for the analysis of daily consumption patterns (Loureiro, 2010).

Table 2. Consumption variables.

Category of time series	Variable	Consumption scenario
Instantaneous flow values	Instantaneous peaking factor (-)	Global, seasonal, weekday
Instantaneous flow values during minimum night consumption period	Average consumption during minimum night consumption period (l/inh·hour)	Global, seasonal, weekday
Instantaneous minimum flow values during night period	Minimum night consumption value (l/service connection·day)	Global, seasonal
Daily flow	Daily peaking factor (-) Average daily consumption (l/inh·day)	Global, seasonal, weekday
Monthly flow	Monthly peaking factor (-) Average monthly consumption l/inh·month)	Global

3. Results

3.1 Case-studies and Data characterization

Data from 2011 were collected from DMA belonging to different Portuguese WDS located in four districts (**Errore. L'origine riferimento non è stata trovata.**): Oporto (*Por*), Braga (*Bra*), Lisbon (*Lis*) and Setúbal (*Set*). Each DMA was identified with a code with an abbreviation of the DMA name and an abbreviation of the district name. For instance, ADE_Bra, refers to a DMA in Braga district.

Loureiro (2010) analysed 22 DMA in the Lisbon district, while in this paper analysed 17 DMA in four different districts. The general characteristics of these DMA are presented in Table 3.

Table 3. General characteristics of analysed DMA.

Characteristic	Interval	Average value	Median value	Loureiro (2010)
Diameter [mm]	75 – 131	100	102	121
Network length [km]	4.0 – 95.0	32	21	19.3
N.º service connections	250 – 3698	1462	1221	643
N.º of clients	742 – 5185	2183	2100	3124
N.º of domestic clients	667 – 4514	1992	1863	2888
N.º of inhabitants	2295 – 9312	3311	2349	5911 (2001)

DMA analysed by Loureiro (2010) have (in average) a lower number of service connections and network length and a higher number of inhabitants, comparing to the DMA analysed in the current work. This is mainly due to the fact that the Lisbon district is characterized by a majority of urban areas with high density of population. This contrasts with more rural areas with lower density of population typical of Oporto and Braga districts.

In order to have a wider understanding of the differences between the northern and southern regions, more data on infrastructure and billing were collected. As a result, a total of 86 DMA were analysed respectively in the north (33) and south regions (53). For these additional DMA, flow time series could not be used due to insufficient historical data or to problems in some data.

The main results in terms of the socio-demographic analysis are depicted in **Errore. L'origine riferimento non è stata trovata.** and show social differences between both analysed regions.

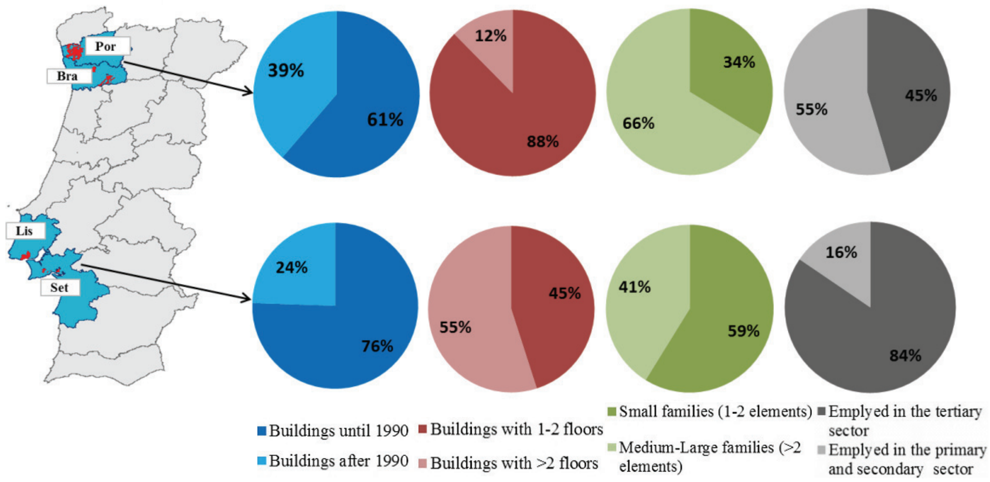

Fig. 2. General overview of socio-demographic variables, at the DMA level.

Socio-demographically, DMA from the north have more recent buildings (39%) than DMA from the south (24%). Buildings with 1-2 floors are more common in the north (88%) than in the south (55%), which corroborates a lower number of service connections and network length in the DMA belonging to the Lisbon district. Concerning the family size, there are only 34% of families with 1-2 elements in DMA from the north, against 59% from the south. In terms of economic mobility, 43% of the workers living in DMA from the north are employed in the tertiary sector of activity, against 81% in DMA from the south. The fact that DMA from the south have a higher percentage of independent workers is probably correlated with higher incomes and may lead to less conservation attitudes towards the use of water (Beal and Stewart, 2011).

Concerning the infrastructure variables, most networks are composed by plastic pipes (>70%). However, the proportion of asbestos cement (AC) pipes is higher in DMA from the south (30%) rather than in the north (3%).

As for the billing variables, the average domestic billed consumption per inhabitant in DMA from the north is 53 l/inh·day, whereas in DMA from the south it is considerably higher with an average of 121 l/inh·day (Fig. 3).

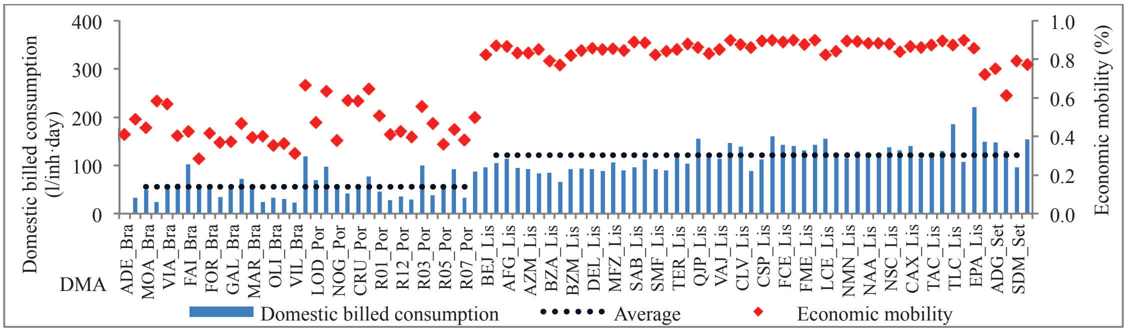

Fig. 3. Domestic billed consumption and economic mobility for the analysed DMA.

As previously referred, the fact that the economic mobility is higher in DMA from the south suggests a higher income, which can be an important factor explaining this difference in the per capita consumption. Additionally, this difference may also be related with the existence of households in the north region that are not connected to the WDS (e.g., households with private wells). The northern region is also characterized by lower temperatures (T) in the summer and much higher precipitation (P) than the south region (i.e., Oporto had an average T= 23.5ºC and an average P=44.5 mm, whereas Lisbon had an average T=26.7ºC and an average T=14.2 mm in the summer of 2011) and thus, outdoor uses in the northern region may be less important. Non-revenue water accounts for 35%, in average, for all DMA analysed with flow data. Public consumption is considerably higher than public: 75% of DMAs have public consumptions below 9.5% and collective consumptions below 3.2%.

In terms of the consumption seasonal scenarios, cluster analysis allowed the distinction of two groups of months. Generally, one group corresponded to the winter season and the other one to the summer season. Fig. 4(a) shows both seasonal scenarios for FAR_Set DMA.

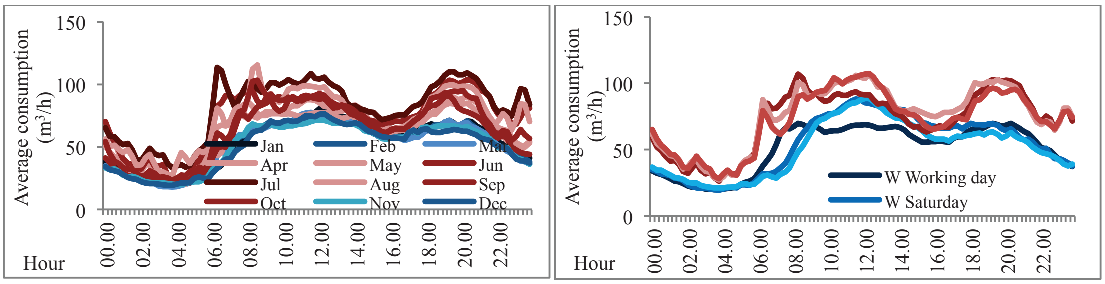

Fig. 4. DMA scenario exploration for FAR_Set DMA: (a) seasonal; (b) weekday.

As depicted, there is an average consumption increase in some months and also a significant behavioural change between January to March (cluster I) and April to September months (cluster II).

Typically for all the analysed DMA, summer scenarios occur from July to September, whereas winter scenarios occur from November to February. Having explored the winter (W) and summer (S) scenarios, the next step was to understand whether the behaviours between days of the week were different for each seasonal scenario. Results showed that working days have a different behaviour from Saturdays and Sundays and bank holidays, for both seasonal scenarios. Thus, for spatial and temporal forecasting three weekday scenarios were analysed for each seasonal scenario. Fig. 4(b) shows the weekday scenario for each season in FAR_Set DMA where significant differences can also be identified.

3.2 Spatial and temporal demand forecasting

3.2.1 Principal Components Analysis (PCA)

Since significant regional differences were identified, PCA was carried out separately for the north and the south region with the aim of finding, from the initial set of 49 variables, the most relevant for each region. For the socio-demographic and infrastructure categories, the first two principal components explained more than 80% of total variance. **Errore. L'origine riferimento non è stata trovata.** summarizes all the variables considered for MLR analysis and describes the structure of each principal component for the south region.

Table 4. Variables considered for MLR with a sample of 53 DMA belonging to the south region.

Category	Predictor variable	Relevant variable (loading)
Socio-demography	PC1: Elderly families	Inactive workers (0.92); Elderly (0.94); Families with 1-2 elements (0.95); Families with 3-4 elements (-0.96);Families with adolescents (-0.97)
	PC2: Individuals mobility	Economic mobility (0.80); University graduates (097); People with 12 years of education (-0.91)
Infrastructure	PC1: Pipe material	Plastic pipes (-0.76); AC pipes (0.78); Service connection density (0.71)
	PC2: Pipe size	% Diameter 110-310 (0.63); % Diameter ≤110 (-0.72)
Billed consumption	PC1: Domestic Billed consumption	Domestic consumption per inhabitant (-0.61); Domestic Consumption 1st level (0.78); Domestic Consumption 2nd level (0.93); Domestic Consumption 4th level (-0.91)
	Commerce and industry billed consumption	Commercial and industrial billed consumption category (%)
	Public billed consumption	Public billed consumption category (%)
	Collective billed consumption	Collective billed consumption category (%)

Concerning the socio-demographic category, the 1st component (PC1: Elderly families) is the most important, as it explains 58.2% of total variance and shows that families with 1-2 elements and inactive workers or elderly are related (positive loadings), in opposition (negative loadings) to families with 3-4 elements and with adolescents. The 2nd component (PC2: Individuals mobility) explains 30.7% of total variance and shows that individuals with higher graduation (university graduates) and working in the tertiary sector (economic mobility) are related, in opposition to individuals with lower education level. For the north region, PCA showed the same components. However, the Individuals Mobility component had a greater importance, since it explained 50.6%, whereas the Elderly families component explained 26.0%.

Regarding the infrastructure, the 1st component (PC1: Pipe material) reflects pipe material, which is independent of network pipe dimension (PC2: Pipe size). In opposition, for the north region, PCA showed that Pipe size is more important than Pipe material.

For billing variables, PCA analysis was only applied to domestic billed consumption variables as they were found to be independent from non-domestic consumption variables. Hence, the only component obtained (PC1: Domestic billed consumption) shows that per capita consumption is related with consumption in higher tariffs. The same structure was obtained when performing PCA in the north region.

Thus, data reduction allowed reducing the 49 initial variables into 8 new variables (5 PCs and 3 variables). A good structure was obtained for both regions and important regional differences were found.

3.2.2 Multiple linear regression for empiric relations

After reducing the variables into components, a correlation matrix was calculated to analyse which relations between consumption (55 dependent variables) and the predictor variables (Table 4) ought to be explored. After analysing the most significant correlations, a MLR analysis was carried out, and some of the most relevant empiric relations obtained are presented in **Errore. L'origine riferimento non è stata trovata.**. Since different behaviours for the north and the south region were detected, MLR was also carried out separately for both regions. In some cases, due to the short number of DMA with flow data in the north region, MLR was only carried out for DMA belonging to the south region.

Table 5. Results of the Multiple Linear regression analysis.

Region	N.º DMA	Dependent variable	Explaining component	Regression coefficient	Standard-Deviation	p-value	r_a^2
North	33	Domestic billed consumption per client (l/cl·day)	Constant (β_0)	318.0	15.5	0.0002	0.53
			Elderly families (β_1)	-29.4	16.7		
			Individuals mobility (β_2)	62.5	17.9		
			Pipe material (β_3)	65.5	16.2		
South	53	Domestic billed consumption per client (l/cl·day)	Constant (β_0)	228.9	5.8	0.0001	0.35
			Elderly families (β_1)	-25.8	6.0		
			Pipe size (β_2)	-13.6	6.0		
	12	Daily peaking factor [-]	Constant (β_0)	1.39	0.04	0.0009	0.81
			Elderly families (β_1)	-0.02	0.03		
			Domestic consumption (β_2)	-0.10	0.03		
			Pipe material (β_3)	0.10	0.04		
		Average consumption per inhabitant (l/inh·day) – winter	Constant (β_0)	172.1	21.3	0.090	0.47
			Individuals mobility (β_1)	34.2	26.7		
			Domestic consumption (β_2)	-21.3	23.7		
		Average consumption per inhabitant (l/inh·day) – summer	Constant (β_0)	220.5	12.1	0.015	0.90
			Individuals mobility (β_1)	34.8	16.6		
			Domestic consumption (β_2)	-55.4	14.7		
		Minimum night consumption per service connection (l/sc·day) – winter	Constant (β_0)	105.5	109.7	0.003	0.73
			Elderly families (β_1)	74.1	42.7		
			Individuals mobility (β_2)	54.0	46.0		
			Commerce and industry consumption (β_3)	28.6	9.9		

In the north region, domestic billed consumption increases more significantly with the Individuals mobility component ($\hat{\beta}_1$=62.5), which means that it is higher for university graduates and individuals employed in the tertiary sector. This regression also indicates consumption is higher for AC pipes ($\hat{\beta}_2$=65.5) and families with adolescents ($\hat{\beta}_3$=-29.4). As previously referred, employees in the tertiary sector may have higher incomes, which may lead to higher water consumption due to less conservation attitudes (Beal and Stewart, 2011).

In the south region, domestic billed consumption is more influenced by the Elderly families component ($\hat{\beta}_2$=-25.8) showing that families with 3-4 elements with adolescents consume more water for domestic purposes. This consumption variable also increases with the Pipe size component ($\hat{\beta}_2$=-13.6), and networks with diameters above 110 mm.

Domestic consumption per client tends to be higher in the north region ($\hat{\beta}_0$=318.0) rather than in the south region ($\hat{\beta}_0$=228.9). This difference is mainly explained by the family size: in the north region, 66% of the families have more than 3 elements, while in the south this statistic accounts for 40%.

Concerning the daily peaking factor, it increases mostly with monthly consumptions above 25 m^3 ($\hat{\beta}_2$=-0.10), plastic pipes ($\hat{\beta}_3$=0.10) and families with 3-4 elements and adolescents ($\hat{\beta}_1$=-0.02). Although not presented in the table, daily peaking factors in the summer increase with public consumption that includes public outdoor uses such as municipal pool fillings and garden irrigation. Zhang (2005) verified that the presence of outdoor uses is an important reason for a significant raise in peaking factors. This is also verified in the monthly and instantaneous peaking factors for all the scenarios analysed. These results support the idea that this type of consumption may have a large potential of improvement if water efficiency measures are taken, especially in the summer period.

In terms of the average flow per inhabitant, two scenarios were analysed: winter and summer. For both scenarios, the average flow per inhabitant is higher for individuals with higher mobility ($\hat{\beta}_1$=34.2) and monthly consumptions above 25m^3 ($\hat{\beta}_2$=-21.3). As expected, consumption tends to be higher in the summer ($\hat{\beta}_0$=220.5) comparatively to the winter ($\hat{\beta}_0$=172.1) (Polebitski and Palmer, 2010).

Concerning the minimum night consumption, it is preferable to analyse this variable in the winter scenario, since the average flows are generally lower for this season and the proportion of consumption due to leakage may be more significant. However, instead of reflecting the infrastructure characteristics that are generally more correlated with leakage results show that this variable is mainly influenced by socio-demographic variables. This may be due the fact that the overall proportion of AC is lower in most DMA (less than 30%) and pipes with this material may be in good conditions. Thus, the minimum night consumption increases with the Elderly families component ($\hat{\beta}_1$=74.1) and the Individuals' mobility component ($\hat{\beta}_2$=54.0). Additionally, it increases with Commerce-industry consumption ($\hat{\beta}_3$=11.1). This tendency also highlights the importance of measuring large consumers for a separate analysis (Loureiro, 2010).

In general, the empiric relations were obtained with better adjustments when comparing to Loureiro (2010), since this study used recently published census data and corresponding to the same time period as all the other collected data. Although the sample used for demand forecasting at DMA level was reduced in some cases, results are encouraging and should be explored with a large number of DMA.

3.2.3 *Correlation analysis for daily consumption patterns*

The analysis of the correlation matrix revealed two different profiles: *young families* and *elderly families*. These results are in agreement with Loureiro (2010). The first includes medium families (3-4 elements) whose elements spend most of the time outside (e.g. work/study outside their municipalities). The second includes small families (1-2 elements), people with more than 65 years or inactive workers and spend most of the time at home. Fig. 5 presents the average consumption patterns for these two consumer profiles for working days in the winter and summer scenarios.

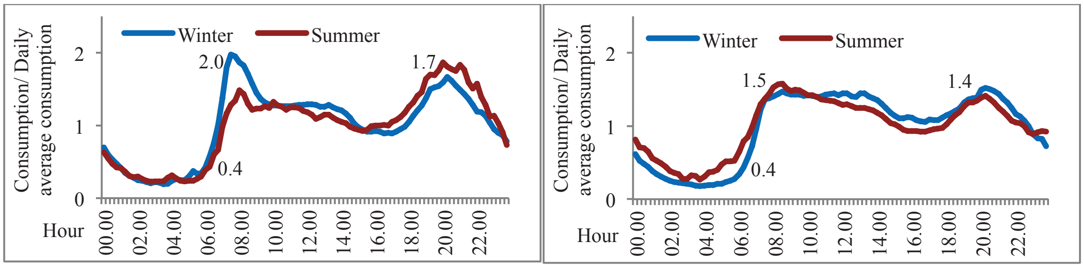

Fig. 5. Average daily consumption patterns for working days (a) *young families*; (b) *elderly families*.

Concerning working days, consumption in the night period (00:00-06:00) is very similar for both young and elderly families, with an average consumption factor of 0.4. *Young families* have a morning consumption peak from 07h30 to 08h30 with a consumption factor of 2.0. On the other side, the consumption pattern for *elderly families* does not show a morning peak, but instead has a flat consumption level from 07h15 to 13h15 with a factor of 1.5. As for consumption during the day (15:00-18:30), both patterns indicate similar consumption, although slightly higher in *elderly families*. Finally, concerning the night peak, the maximum consumption is registered at 20h15 with an average consumption factor of 1.7 for *young families* and 1.4 for elder families. No significant differences between winter and summer scenarios were observed, apart from the morning peak being higher during the winter season for *young families* and minimum flows in the night period being lower in the same season for elderly families.

4. Conclusions

The goal of the current work was to combine a large amount of data from different sources and resolutions to develop a spatial and temporal forecasting of water consumption and patterns in WDS.

A consolidated and systematic methodology was adopted and a framework of 86 DMA with 49 socio-demographic, billing and infrastructure variables was developed. Scenario exploration over flow time series was carried out and allowed the identification of seasonal (i.e., winter and summer) and weekday scenarios (i.e., working days, Saturdays and Sundays). Using these trends, 55 consumption variables and 6 daily consumption patterns were analysed.

A rigorous Principal Components Analysis (PCA) was carried out to define the most relevant socio-demographic, billing and infrastructure variables, followed by correlation and Multiple Linear Regression (MLR) analyses. The most important components behind the empiric relations obtained are socio-demographic and indicate that consumption is higher for large families employed in the tertiary sector (higher incomes), families with adolescents and university graduates. Public consumption is also relevant, especially to explain consumption increase in the summer season.

For estimating consumption patterns, two different consumption profiles have been identified (i.e. *young and elder families*), clearly showing that different daily consumption behaviours are mainly associated with different family structures (i.e. families with adolescents or elderly).

Although the number of cases used for demand forecasting at DMA level was reduced in some cases, results are encouraging and should be explored with a large number of DMA. This contribution considerably reduces the uncertainty in planning and operation of water distribution systems. Further developments include testing the empiric relations obtained in other DMAs with flow data.

Aknowledgements

The authors would like to thank Rita Almeida from AGS, S.A, Catarina Sousa from Águas do Sado, S.A. and André Pina from SMAS de Oeiras e Amadora water utilities who contributed to this work by providing their data.

References

Beal, C., Stewart, R.A., 2011. South East Queensland Residential End Use Study: Final Report, in: Alliance, U.W.S.R. (Ed.). Griffith University.

Donkor, E., Mazzuchi, T., Soyer, R., Roberson, J., 2012. Urban Water Demand Forecasting: A Review of Methods and Models. Journal of Water Resources Planning and Management.

Farley, M., Trow, S., 2003. Losses in water distribution networks. A practitioner's guide to assessment, monitoring and control. IWA Publishing, London.

Grafton, R.Q., Ward, M.B., To, H., Kompas, T., 2011. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. Water Resources Research 47.

INE, 2012. Census 2011: Final results - Portugal. Instituto Nacional de Estatística - Statistics Portugal, Lisbon.

IRAR, 2009. Tariff formation for end-users of drinking water supply, urban wastewater and municipal waste management services. ERSAR.

Loureiro, D., 2010. Consumption analysis methodologies for the efficient management of water distribution systems (*in portuguese*). Technical University of Lisbon, Lisbon, Portugal.

Mamade, A., 2013. Profiling consumption patterns using extensive measurements - A spatial and temporal forecasting approach for water distribution systems. Instituto Superior Técnico, Lisbon.

March Corbella, H., Saurí i Pujol, D., 2009. What lies behind domestic water use?: a review essay on the drivers of domestic water consumption. Boletín de la Asociación de Geógrafos Españoles, 297-314.

Polebitski, A.S., Palmer, R.N., 2010. Seasonal Residential Water Demand Forecasting for Census Tracts. Journal of Water Resources Planning and Management 136, 27-36.

Rebelo, M., Freitas, M.J., 2004. Socio-urban analysis and interpretation of Lisbon - Progress report 2. National Laboratory for Civil Engineering, Lisbon.

Shandas, V., Hossein Parandvash, G., 2010. Integrating urban form and demographics in water-demand management: an empirical case study

of Portland, Oregon. Environment and planning. B, Planning & design 37, 112.
Zhang, X., 2005. Estimating Peaking Factors with Poisson Rectangular Pulse Model and Extreme Value Theory. University of Cincinnati.