

Modeling the effect of weather conditions on urban water demand in multiple network areas: a practical approach to improve monthly and seasonal operation

Marta Cabral, Aisha Mamade, Dália Loureiro, Conceição Amado and Dídia Covas

ABSTRACT

The current paper is aimed at studying the influence of weather conditions on urban water consumption using available continuous flow measurements from existing district metered areas (DMA), belonging to different Portuguese water distribution systems. A three-step methodology was followed: Data processing; DMA segmentation in terms of billed consumption and socio-demographic characteristics; and demand modeling. Cluster analysis was carried out to segment the different DMA with similar billed consumption, and multiple linear regression analysis was used to describe the effect of weather variables (temperature and rainfall) on water consumption. Three well-defined DMA clusters were identified and the respective demand models were obtained for each group. Results have shown that temperature has a significant influence on water consumption, particularly in DMA in which the public billed consumption (29%) and the domestic billed consumption in the third and fourth tariff levels (10.7 and 10.1%, respectively) are relevant. The relationship between consumption and rainfall is not as evident as the relationship with temperature; however, the rainfall variable is equally important in the construction of demand models. These results allow improving of monthly and seasonal network operation, since these allow a more robust profiling of the water uses as well as modeling demand for different weather conditions.

Key words | demand modeling, district metered areas, regression analysis, water distribution systems, weather conditions

Marta Cabral (corresponding author)
Aisha Mamade
Dália Loureiro
Urban Water Unit,
National Civil Engineering Laboratory,
Lisbon,
Portugal
E-mail: mfcabral@lnec.pt

Aisha Mamade
Dídia Covas
CERIS, Instituto Superior Técnico,
Universidade de Lisboa,
Lisbon,
Portugal

Conceição Amado
CEMAT, Instituto Superior Técnico,
Universidade de Lisboa,
Lisbon,
Portugal

INTRODUCTION

Water is a natural resource of utmost importance to the prosperity and development of society, with a crucial role in the location and development of communities (OECD 1999). Water management aims at providing the adequate amount of water with adequate quality for various water uses. Particularly in developed societies, water systems have been modified or adapted to fulfill the demands of the water services in the long term (Haasnoot *et al.* 2011).

The understanding of domestic consumption—the most important component of urban water consumption—is important for adequate management of existing water

supply systems, for the design of new systems or the expansion of existing sectors, as well as for the establishment of long-term strategies (Silva *et al.* 1995; Rebelo *et al.* 2008; Pullinger *et al.* 2013). Overall, knowledge of the factors that influence domestic consumption is important for efficient planning, operation and management of water supply systems. Domestic water consumption is strongly influenced by economic, socio-demographic and weather conditions, infrastructure characteristics, technological developments and regulation factors (Höglund 1999; Arbués *et al.* 2004; Corbella & Pujol 2009; Loureiro 2010). Weather conditions

are one of the most relevant explanatory factors of domestic water consumption (González & Carranza 2003). Temperature and rainfall are the most commonly used weather indicators, given the availability of data. Miaou (1990) suggest that water use can be decomposed into a base use and a seasonal use; the former is primarily indoor use and is not affected by weather conditions, whereas the latter is largely outdoor use and is strongly dependent on weather conditions.

Temperature is a key factor for domestic water consumption, particularly in networks with significant outdoor uses (Cabral 2014). Typically, hotter days are associated with higher water consumption through increased hygiene habits and outdoor uses, such as garden watering and swimming pool filling (Hoffmann *et al.* 2006; Corbella & Pujol 2009). Loh & Coghlan (2003) argued that consumption in outdoor uses is very sensitive to weather conditions. These authors verified that low-rise buildings with high income inhabitants have a higher consumption in the summer, mainly due to outdoor uses.

Miaou (1990) suggested that the effect of rainfall on daily urban water use is both dynamic and state-dependent. Dynamic means that the occurrence of rainfall causes a temporary reduction in seasonal water use that diminishes over time and, eventually, becomes negligible. State-dependent implies that, under the same rainfall conditions, the higher the seasonal water use level is (prior to the occurrence of a rainfall), the greater the rainfall effect that can be expected. The rainfall state-dependent effect has two important implications. Firstly, people respond more to its occurrence than to its amount; in other words, the effect is more psychological than physical (at least in the short-term). Secondly, the rainfall has hardly any effect on consumption when water use is mainly due to indoor uses, which is either a result of low temperature in the winter or of several days of consecutive rainfall. Parandvash & Chang (2016) suggest that precipitation has an impact on daily water consumption in a statistically significant way.

Knowledge of the influence of weather conditions on water consumption is useful to support short, medium and long-term decisions for the efficient use of water and the sustainable management of urban water systems. It is important to understand in which network sectors the urban consumption is more sensitive to weather conditions. For example,

the consumption in areas with relevant outdoor uses (e.g. gardens, pools) or in tourist attractions may be more influenced by weather conditions (Gössling *et al.* 2012). This aspect is important to improve the operational management of the network sectors, given the increase of consumption scenarios for increasing temperature.

Additionally, in order to improve water demand management in situations of scarcity and to be able to apply measures to promote the efficient use of water, it is necessary to better understand which water uses are more influenced by weather conditions. This analysis is particularly important in the context of climate change. The increase of the average annual temperature, the changes in spatial and temporal distribution of rainfall and the variations in the frequency and intensity of extreme weather events can generate constraints on water availability and changes in water use (Easterling *et al.* 2000).

Weather conditions have been used in the construction of water demand models in the previous research. Miaou (1990) proposed a new class of urban water demand model with nonlinear climatic effect based on monthly time-series. Martinez-Espineira (2002) calculated and compared short- and long-run price elasticities of residential water demand using annual time-series from Seville, Spain. This author included in the demand model two variables, the average daily maximum temperatures of each month and current level of rainfall, and has taken into consideration the co-integration and error correlation techniques. Babel *et al.* (2007) used the annual time-series of Kathmandu Valley in Nepal to develop a linear, semi-log and log-log models, including the average annual temperature and annual rainfall as explanatory variables.

One of the most important methods to improve network operation is to divide the water distribution system into smaller network units, called district metered areas (DMA), with 500–3,000 service connections, and to install flowmeters to measure the inflow and outflow from these areas (Farley & Trow 2003). Most of the previous studies have used monthly or annual time-series to construct the demand models, considering a larger spatial scale (e.g. city level). Flow time-series with high resolution (15 minutes) at DMA level are used herein. This is important to support network operation (e.g. pump scheduling, tank filling), as the obtained results allow identifying and quantifying in

which DMA the influence of weather conditions on urban water consumption is more relevant.

The current research aims to improve the understanding of the influence of weather conditions on water consumption using available flow measurements of Portuguese DMA located in different water distribution systems (WDS). In this study, the term *consumption* refers to all components of urban water consumption (i.e. domestic, non-domestic and leakage), whereas the term *flow data* solely refers to flow measurements in the network. A three-step methodology was applied to determine the DMA clusters and to develop the respective demand models between water consumption and weather variables (i.e. temperature and rainfall).

The main contributions of this research are: (i) the development of an improved methodology to describe the effect of weather conditions on water consumption, using cluster analysis (CA) to segment DMA in terms of billed consumption, obtaining different clusters according to water uses; (ii) the application of this methodology at the operational level (DMA level) and using time-series with high resolution (15 minutes); and (iii) the development, testing and validation of novel demand models between water consumption and weather conditions, taking into consideration the serial correlation of error terms, the autocorrelation and cross-correlation between flow time-series and the logarithm transformation of the models. Results of obtained models contribute to improving the daily operation of the network areas, since these allow profiling of the water uses of each DMA and modeling of water demand for different scenarios of temperature and rainfall. Although the demand models are case-specific, they can be applied to sectors with billed consumption and socio-demographic characteristics similar to the analyzed ones.

METHODOLOGY

This research follows a three-step methodology to study the influence of weather conditions on water consumption: (1) data processing, (2) DMA segmentation in terms of billed consumption and socio-demographic characteristics, and (3) demand modeling, taking into consideration weather variables for each DMA group (Figure 1). Figure 1 also shows details of temporal and spatial scales and data

source for the different types of data. This methodology is different from the usual approaches that forecast the water demand (see, for example, House-Peters & Chang (2011), Donkor *et al.* (2012), Behboudian *et al.* (2014) and Lee & Chae (2015)) using different explanatory variables of consumption, but it allows the identification of the network sectors in which the consumption is more sensitive to weather conditions (temperature and rainfall) and quantifying this influence.

In Step 1, available flow measurements from each DMA are collected with either an irregular or regular time-step over one year. The used data processing methodology is the one proposed by Loureiro (2010). Network flow data processing includes descriptive analysis, outlier detection, data combination and data normalization. Descriptive analysis aims at data characterization and at the identification of potential outliers or of less reliable data (Loureiro *et al.* 2015). Data combination includes removing the large consumers and the minimum night consumption (usually associated to water losses). Data normalization consists of converting the flow series into a regular time-step of 15 minutes. Billing, weather and socio-demographic data are also collected with different temporal and spatial scales. This data processing allows representative data at the DMA level to be obtained.

In Step 2, a CA is carried out to segment the different DMA in which the billed consumption characteristics are similar, as well as their type of water uses. Ward's method (Ward 1963) and Euclidean distances with standardized variables are used herein in CA. In this hierarchical method, the number of clusters is defined after the clustering process, based on the linkage distance.

Step 3, demand modeling, allows the identification of the influence of weather conditions on the spatial (i.e. network sectors) and temporal distribution of consumption (i.e. seasonal variations in consumption). Multiple linear regression (MLR) analysis of consumption variables of each cluster is carried out, as a function of temperature and of rainfall:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

where Y is the dependent variable (e.g. consumption); x_i is the explanatory variable (e.g. temperature or rainfall);

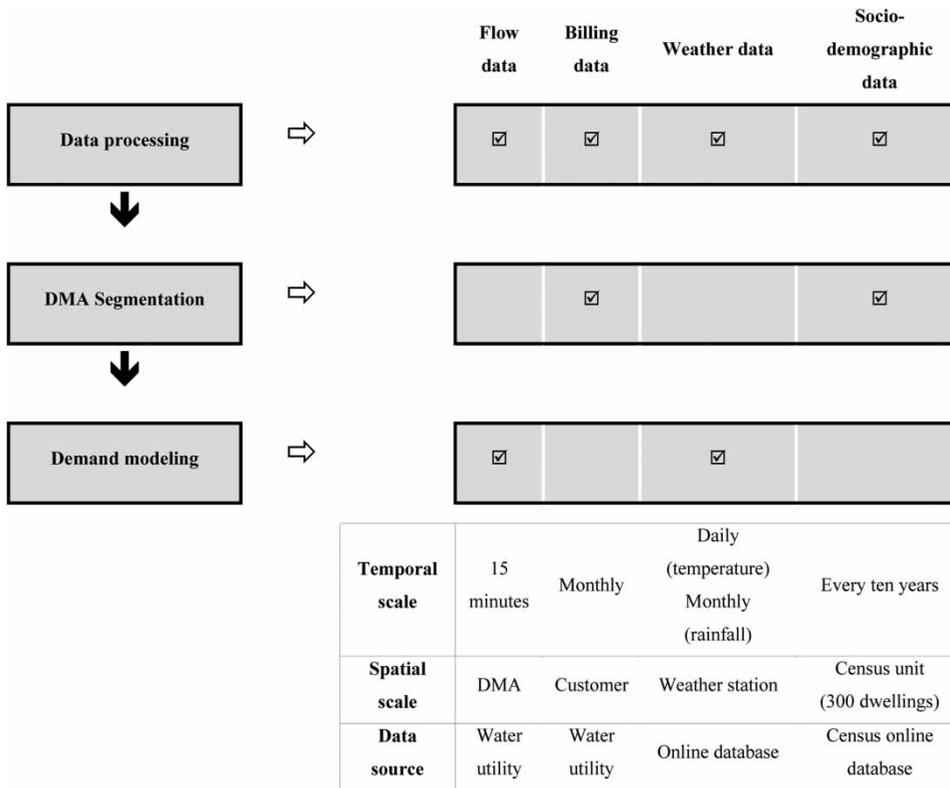


Figure 1 | Adopted methodology to study the influence of weather conditions on water consumption, type of data used in each step and data characterization.

β_0 are the regression parameters; β_0 is the constant term; β_1, \dots, β_p are the regression coefficients and represent the increase (positive value) or the decrease (negative value) of the dependent variable associated with a unit variation of the explanatory variable; and ε is the random component that represents the disturbance or error term. This regression model helps to assess the existence of relationships between the variables Y and x_1, \dots, x_p .

Different measures of goodness-of-fit should be used to evaluate the fitted regression models, namely, standard deviation/error of the estimator's beta and adjusted r-square (r_a^2). The r_a^2 is a modification of r-square adjusted to the number of existing explanatory variables. It represents the quality of the adjustment: a value close to 1 indicates that the adjustment of the regression is very good and the linear regression can explain most of the variation in the dependent variables. The p -value associated to the overall F-test is a measure of model significance, wherein the null hypothesis $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ is rejected when the p -value is, for example, less than 0.05.

The variance inflation factor (VIF) is a measure of the degree of multi-collinearity between explanatory variables in a multiple regression model (O'Brien 2007). Multi-collinearity can inflate the variance amongst the variables in the model. These inflated variances are problematic in regression, because some variables add very little or even not new and independent information to the model. A general rule is that the VIF should not exceed 10 (Robinson & Schumacker 2009). This measure is also used in Step 3.

The autocorrelation and cross-correlation of time-series and the serial correlation of error terms are taken into account in the development of the MLR models. The autocorrelation function (ACF) is a measure of how a quantity observed at a given time is related to the same quantity at another time. It measures the degree of resemblance of the signal with itself as time passes, to use signal processing terminology, while the cross-correlation function (CCF) measures how closely two different observations are related to each other at the same or different times (Scargle 1989). In the relationship

Table 1 | Criteria used for DMA selection in terms of billed consumption and economic characteristics

Billed and economic parameters	Description	Criteria
Domestic billed consumption (%)	Ratio between annual domestic billed consumption and the total annual billed consumption.	>50%
Domestic billed consumption in the 3rd and 4th tariff levels (%)	Ratio between annual domestic billed consumption in the 3rd and 4th tariff levels and the total annual domestic billed consumption in all tariff levels (1st to 4th tariff levels). According to IRAR (2009), the 3rd tariff level is defined between 15 to 25 m ³ /month and the 4th tariff level is defined for volumes higher than 25 m ³ /month.	>5%
Public billed consumption (%)	Ratio between annual public billed consumption and the total annual billed consumption.	>5%
Economic mobility index (%)	Ratio between the number of residents employed in the tertiary sector and the number of total residents. Previous theoretical models used the income variable to predict residential water demands. Cole (2004) indicates that higher levels of income can indicate higher living standards, which could imply a higher quantity of water consumption.	>70%

between two time-series (Y_i and x_i), the series Y_i may be related to past lags of the x-series (Haugh 2010).

The study of the serial correlation of error terms is carried out through the Durbin–Watson test and global validation of linear model assumptions. The Durbin–Watson test computes residual autocorrelations, allowing the independence of error terms to be concluded (Durbin & Watson 1950). The global validation of linear model assumptions included the global stat, skewness, kurtosis, link function and heteroscedasticity tests. All tests show the bootstrapped p -values, in which the null hypothesis is not rejected when the p -value is higher than 0.05.

Case study

Analyzed case-studies include DMA from two different WDS from Lisbon and Setubal districts in Portugal. To identify the subset of DMA where temperature and rainfall may have a higher influence on urban water consumption, billing and socio-demographic variables were explored for each DMA according to the criteria described in Table 1. Billing variables include total domestic billed consumption, domestic billed consumption in the third and fourth tariff levels and total public billed consumption. DMA were selected in order to ensure that the main component of urban consumption (Almeida *et al.* 2006) – domestic – is predominant in these network areas. An additional criterion was considered in terms of the proportion of domestic consumption in the higher tariff levels in order

to focus the analysis on areas with higher consumption per customer.

The public billed consumption includes outdoor water uses in fountains, street washing, irrigation and cleaning of sewers (Almeida *et al.* 2006). In this study, this variable also includes consumption in public institutions (e.g. schools, sporting installations), since outdoor uses of these customers (e.g. gardening and swimming pool operation) might also be significant.

Moreover, an additional socio-demographic criterion related to residents that work in the tertiary sector was considered. Workers in the tertiary sector are usually associated with higher incomes, which can be related to a higher consumption and often less conservation attitudes towards the efficient use of water (Loh & Coghlan 2003; Beal & Stewart 2011). For the calculation of the economic mobility index, data from the last census were collected (INE 2012).

The previous criteria were applied to an original set of 44 DMA and allowed to select a total number of 10 DMA. These network sectors are located in the south region of Portugal: seven correspond to the district of Lisbon, with the flow series of the years 2006 and 2007, and the remaining three DMA correspond to the district of Setubal, with the flow series of the year 2011. The general characteristics of analyzed DMA are presented in Figure 2, using box-plots. This schematic representation allows the describing of the data variation through their quartiles, including the minimum, first quartile, median, third quartile and maximum.

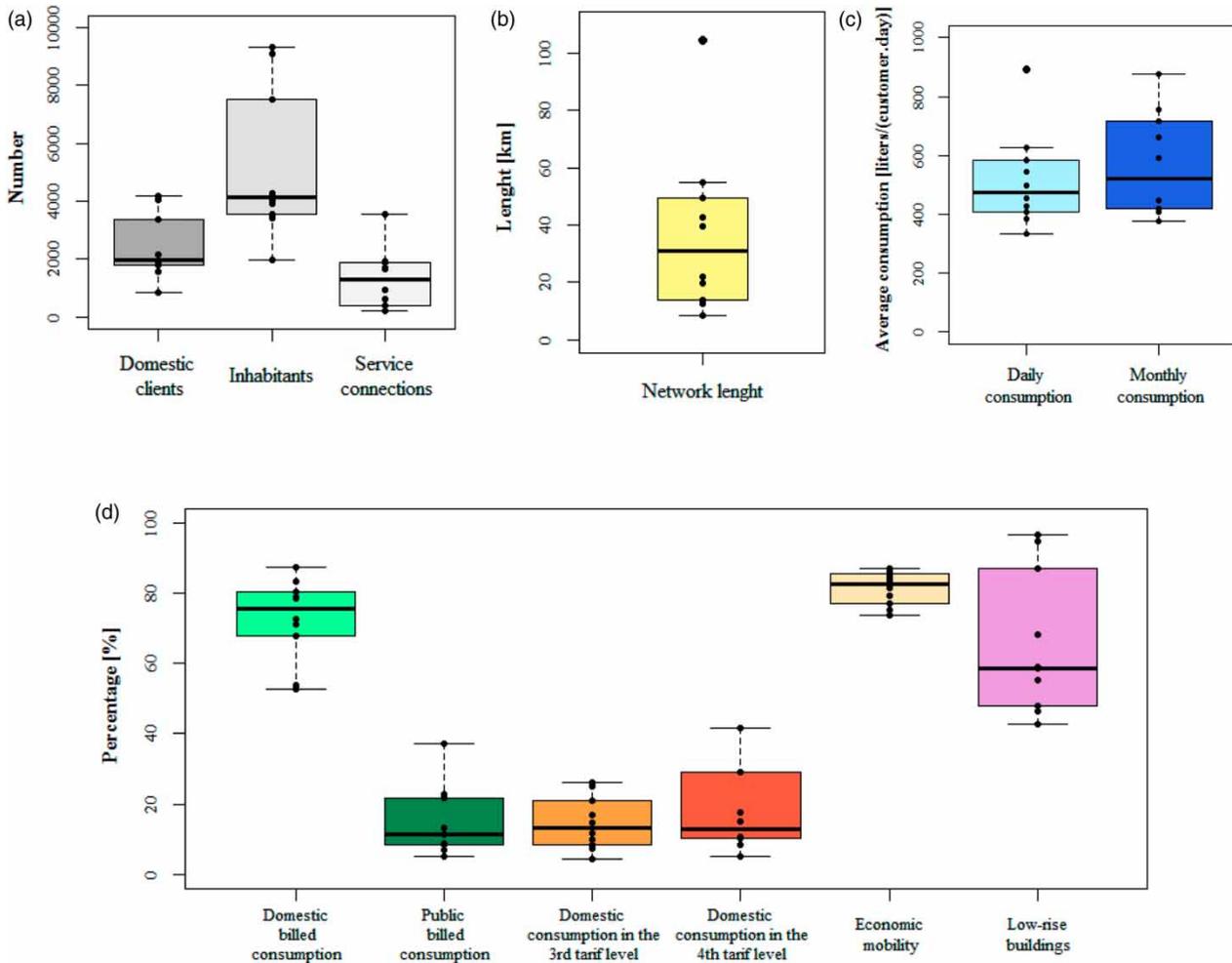


Figure 2 | General characteristics of analyzed DMA: (a) network parameters, (b) network length, (c) consumption parameters, and (d) billing and socio-demographic parameters.

The influence of temperature and rainfall on water consumption was studied through the average daily temperature and monthly rainfall as explanatory variables. Temperature data were collected from the available weather database (Weather Underground 2015), which provides the average daily temperature measured at different weather stations. Rainfall data were collected from the Portuguese National Information System of Water Resources (SNIRH 2015), which provides only the monthly rainfall measured at different weather stations in Portugal.

For the Lisbon DMA, the weather stations of Lisbon airport and S. Julião do Tojal for 2006–2007 were used to collect temperature and rainfall data, respectively. For the Setúbal DMA, the average daily temperatures from the weather station of Várzea for 2011 were collected and the

monthly rainfall data were collected from the weather station of Moinhola. Table 2 presents the variation of temperature and rainfall for the different DMA and considered years. Average national temperatures are also presented, indicating an annual temperature for the studied DMA, lower compared with the average national temperature. Figure 3 depicts the average monthly temperature and monthly rainfall variation for the one-year period of analysis: temperature varies between 10 and 25 °C and the rainiest year was 2006, with a higher incidence of rainfall in March, October and November.

Each DMA presents different values of consumption and economic parameters (domestic billed consumption, public billed consumption, domestic billed consumption in the third and fourth tariff levels, and economic mobility

Table 2 | General overview of climate variables (average monthly temperature and monthly rainfall)

District	Year	Temperature (°C)			Rainfall (mm)			Average national temperature (°C)	Year
		High	Average	Low	High	Average	Low		
Lisbon	2006	25	18	10	2006	95	2	16	2006
Lisbon	2007	23	17	11	2007	36	3	15	2007
Setúbal	2011	25	19	13	2011	57	1	16	2011

index); these values belong to different years. The DMA located in Lisbon has flow time-series from 2006 and 2007, however the consumption and economic parameters refer to 2011. In the DMA located in Setúbal the flow time-series and the consumption and economic parameters are from the same year, 2011. Only DMA from well-established urban areas were used, in which changes in terms of buildings, population and family are negligible, allowing the use of the flow time-series and the parameters with different years between them.

RESULTS

The consumption parameters of each DMA (billed consumption and consumption tariff levels) were used for the CA. Analyzed DMA show high percentages of economic mobility (>70%) with workers in the tertiary sector, although with low variability between them (74–87%) and, therefore, this parameter was not included in the CA.

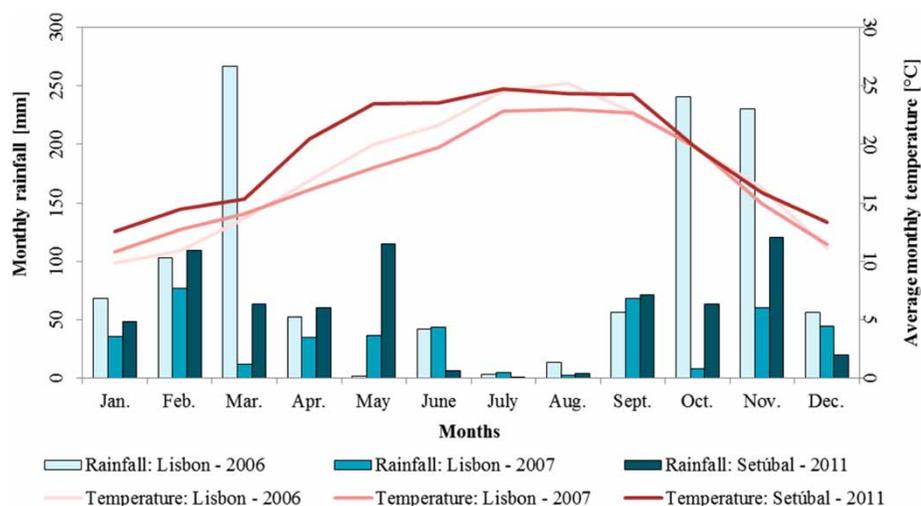
Three well-defined clusters were identified, considering a linkage distance equal to 50, as shown in Figure 4. Cluster 1 consists of two DMA from Lisbon; Cluster 2 comprises three DMA from Setúbal; and Cluster 3 consists of five DMA from Lisbon.

Two consumption variables were analyzed to understand the main differences between the clusters identified, N being the number of observations of each cluster:

- average monthly consumption ($L/(\text{customer.month})$) considering the annual flow time-series and all weekdays (Figure 5);
- average daily consumption ($L/(\text{customer.day})$) considering the annual flow time-series and only the working days (Figure 6).

The general characteristics of the analyzed clusters are presented in Table 3.

Results have shown that consumption increases with the temperature and decreases with the rainfall (Figure 5).

**Figure 3** | Average monthly temperature and monthly rainfall.

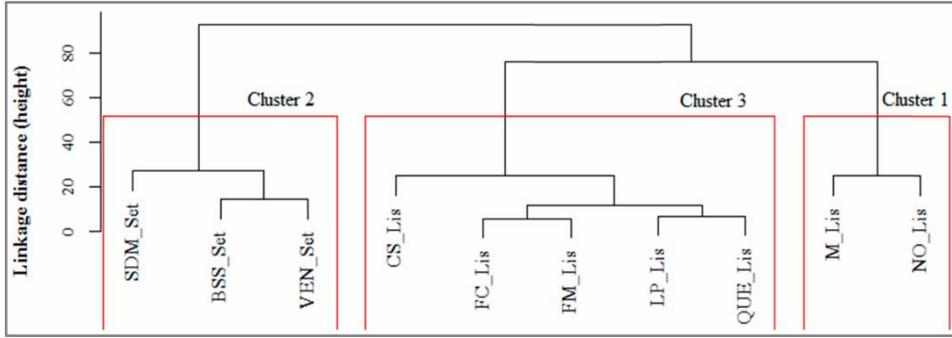


Figure 4 | Hierarchical clustering dendrogram of the DMA using consumption billing variables.

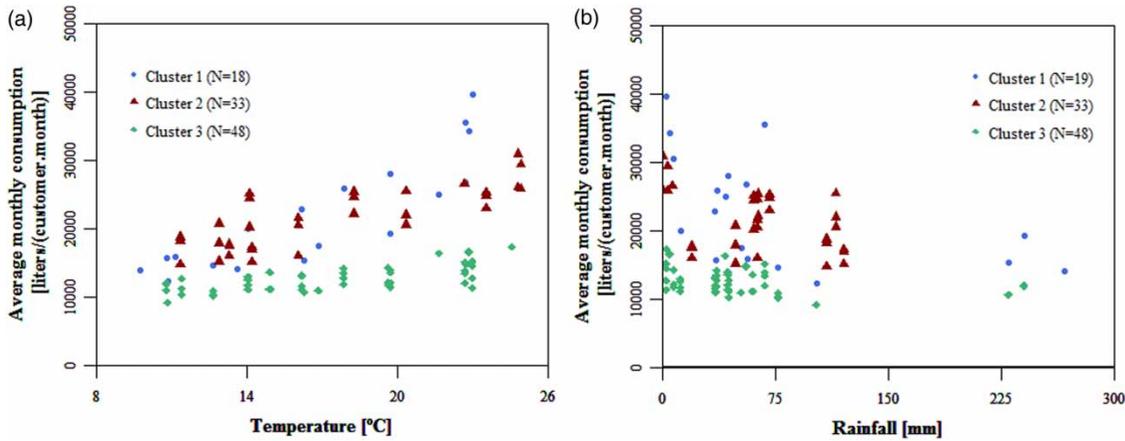


Figure 5 | Representation of average monthly consumption for the clusters identified: (a) with the temperature and (b) with the rainfall.

The DMA consumption in Cluster 1 is the most affected by the temperature. This cluster is composed of DMA with the highest public billed consumption (almost 30% of total

billed consumption) and the lowest domestic billed consumption (ca. 50%). This shows the importance of temperature on non-domestic consumption (mainly, public consumption). Cluster 2 has the largest domestic billed consumption in the third and fourth tariff levels (around 67%); these consumptions are related to outdoor uses during the summer period. Cluster 3 has a low temperature influence on water consumption and is composed of the largest number of DMA. The relationship between consumption and rainfall is not as evident as the relationship between

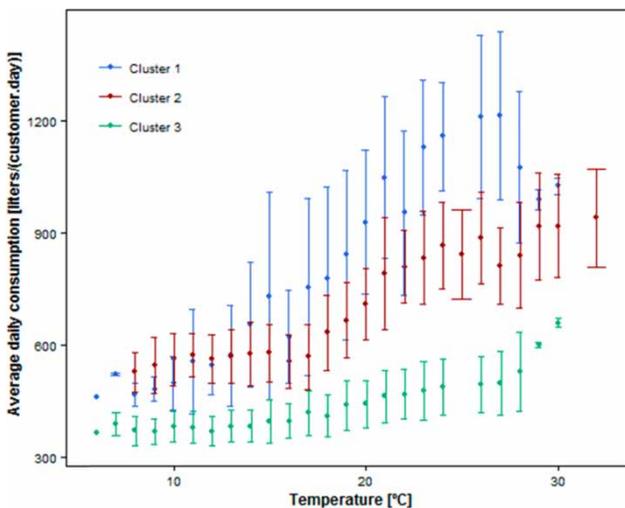


Figure 6 | Average daily consumption error bars for each value of temperature.

Table 3 | General characteristics of analyzed clusters

	Domestic billed consumption (%)	Domestic billed consumption in the third and fourth tariff levels (%)	Public billed consumption (%)
Cluster 1	53	21	29
Cluster 2	80	67	11
Cluster 3	78	20	8

consumption and temperature (for the three clusters), presenting a larger dispersion of data.

The average daily consumption was only studied as a function of the temperature due to the data availability for this temporal scale (Figure 6). Consumption increases with the temperature, and Cluster 1 is the most affected by temperature. Error bars representing the standard deviation of average daily consumption for the different values of temperature were graphically plotted for the three clusters. Additionally, for each cluster, the variability of consumption increases with the average daily consumption, with the highest variability being for Cluster 1.

In order to characterize the daily consumption behavior of each cluster, instantaneous (15 minutes) dimensionless consumption patterns were obtained (Vitorino *et al.* 2014). This analysis allows the association of the consumption patterns for water uses through defined clusters. Three types of daily consumption patterns were obtained and are presented

in Figure 7, each corresponding to a cluster. Although these patterns include all types of consumption (i.e. domestic, non-domestic and leakage), their daily variation reflects the domestic consumption, which is higher than 50% of the total consumption. This analysis considered only the working days and the winter period, due to a more homogenous consumption in this period of the year.

Clusters were obtained using the consumption parameters of each DMA (billed consumption and consumption tariff levels) and the obtained consumption patterns are similar in each cluster, indicating that the chosen parameters are explanatory of consumption. Cluster 1 reflects a non-domestic pattern due to the high public consumption. This fact contributes to a higher consumption factor in the night period (from 0100 to 0600 h) compared with the other clusters, mainly due to the watering of gardens usually carried out during this period. Clusters 2 and 3 present similar daily domestic patterns, with significant consumption factors in the morning

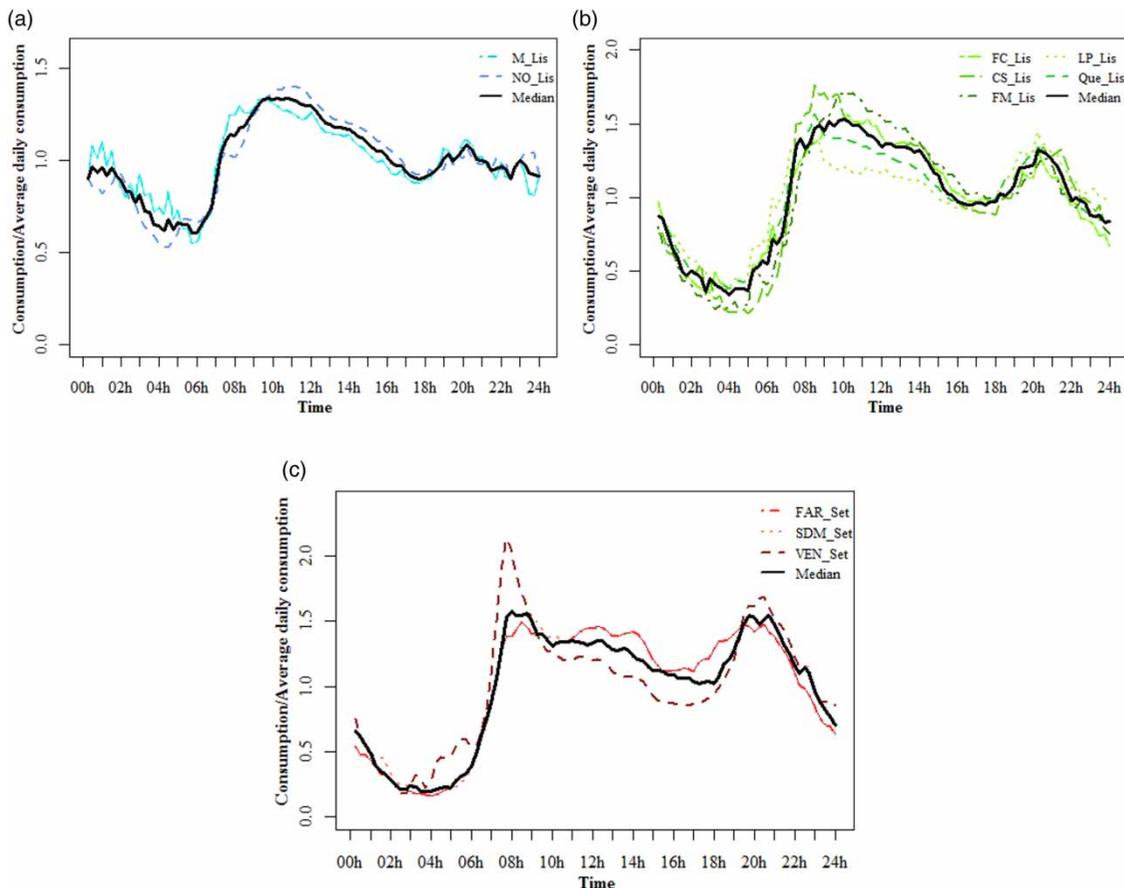


Figure 7 | Dimensionless daily consumption patterns for the three clusters considering only the winter period and working days: (a) Cluster 1, (b) Cluster 2, and (c) Cluster 3.

period (from 0700 to 1000 h) and in the dinner period (from 1900 to 2200 h). Cluster 2 presents higher consumptions in these periods and the morning period occurs earlier than in Cluster 3; these consumption patterns are usually from areas where individuals spend most of the time outside home (e.g. at work or school outside their municipalities). In all clusters the median consumption patterns were represented.

The variation of average monthly consumption throughout the year for the three clusters is presented in Figure 8. Cluster 1 presents the highest average consumption in the summer period in comparison with the other clusters. This result suggests the importance of non-domestic consumption in Cluster 1, since public consumption contributes to increased watering of gardens, with the highest incidence in the summer period. Cluster 2 is composed of DMA with larger domestic billed consumption in the third and fourth tariff levels associated to outdoor uses, such as swimming pools, contributing to an increase of consumption in the summer period. Cluster 3 presents the lowest variation of average monthly consumption throughout the year, which may be explained by the high percentage of domestic consumption in the first tariff level associated with indoor uses (water base use), which is less sensitive to weather conditions.

Since Cluster 1 had the highest influence of weather conditions on water consumption, the average daily consumption variation throughout the year and the seasonal factor considering average annual consumptions were further analyzed. The average daily consumption variation is represented by the box-plot (Figure 9), in which August has the highest consumption with a median value of 1,316 L/customer.day. The seasonal factor is represented by the points with a seasonal

factor variation between 0.63 and 1.58, respectively, in January and August (Figure 9). The high values of the seasonal factors are associated with public consumption, with a higher incidence in the summer periods.

Three clusters were identified by CA, and this was followed by the construction of MLR models for each cluster. Obtained statistics and MLR results are presented in Table 4 for average monthly consumption (Q_m). Presented MLR models were the best-fitted ones after autocorrelation and cross-correlation analysis, and serial correlation of error term tests.

Autocorrelation analysis was developed for the dependent variable (average monthly consumption) for each cluster through the ACF function that analyzes the different consumption lags. It was found that consumption in the current month is dependent on the previous month's consumption in all clusters, and it is represented by the variable C_{-1} in Table 4. The cross-correlation analysis was developed between the dependent variable (average monthly consumption) and each explanatory variable (temperature and rainfall). The CCF function allows the lags (different month's times) of explanatory variables that are more correlated to the dependent variables to be obtained. It was found that consumption in the current month is dependent on the previous month's temperature, only in Cluster 2, and it is represented by the variable T_{-1} in Table 4. The negative value of the T_{-1} variable means that the temperature of the previous month has a negative effect on the monthly consumption. A higher temperature in the previous month will contribute to a lower consumption in the current month, for example in the months of August and September. On the other hand, a lower temperature in the previous month

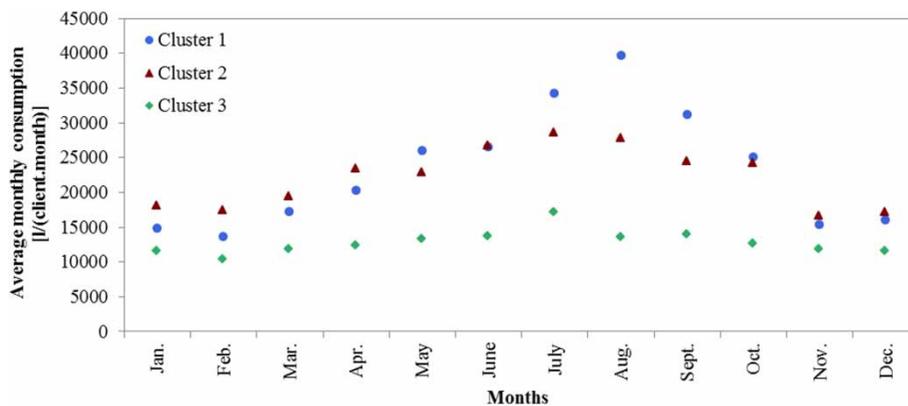


Figure 8 | Variation of the average monthly consumption throughout the year for the three clusters.

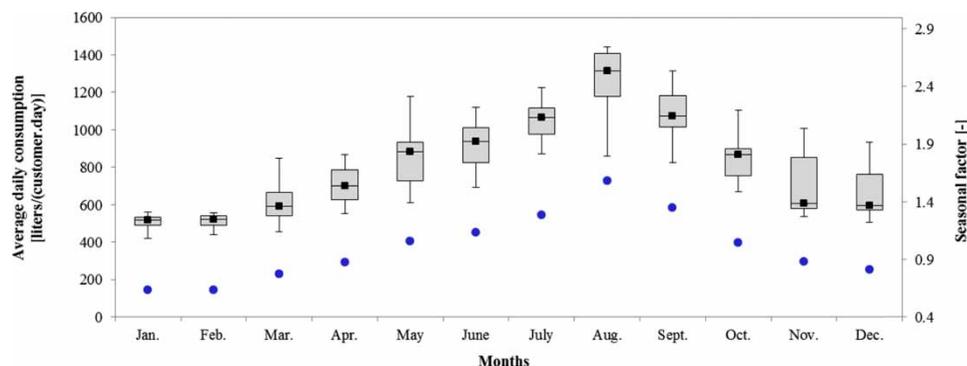


Figure 9 | Average variation of the daily consumption throughout the years of 2006 and 2007 in each month and seasonal factor for Cluster 1.

will contribute to a higher consumption in the current month, for example in the months of May and June. Besides the use of C_{-1} and T_{-1} (only in Cluster 2), the variables of rainfall and temperature corresponding to the current month were also used as explanatory variables for all clusters.

A minor logarithmic relationship of the consumption data with temperature was observed for Cluster 1 (see Figure 5(a)); thus, MLR models with logarithmic transformation were constructed for this dependent variable (average monthly consumption). In the other clusters, the best-fitted models corresponded to MLR models without logarithmic transformation.

The study of the serial correlation of error terms was carried out through the Durbin–Watson test and the global validation of linear model assumptions. The p -value obtained in the Durbin–Watson test is also presented in

Table 4; for the three models, the p -value is higher than 0.05, which corresponds to the value at which the null hypothesis is not rejected. Also, in all clusters, the assumptions concerning the global validation of the linear model were satisfied, including the assumptions of global stat, skewness, kurtosis, link function and heteroscedasticity.

Results of the MLR analysis of average monthly consumption (Q_m) present r_a^2 values relatively high (≥ 0.60) and approximately null the p -values (overall F-test) for the three clusters. Ninety-three percent of the variations of average monthly consumption in Cluster 1 are explained by the independent variables, since $r_a^2 = 0.93$. Consumption in Cluster 3 is characterized by the lowest adjusted r-square ($r_a^2 = 0.60$), being the cluster with the highest percentage of domestic consumption in the first tariff level associated with indoor uses (water base use), which is very insensitive to weather

Table 4 | MLR models for average monthly consumption

Dependent variable	Clusters	Explaining component	Regression coefficients	Standard deviation	p -value (F-test)	r_a^2	p -value (Durbin–Watson test)
Q_m , average monthly consumption [l/(customer.month)]	Cluster 1 ^a (N = 18)	Constant (β_0)	5.632	1.071	0.000	0.93	0.962
		C_{-1} (β_1)	0.378	0.120			
		R_0 (β_2)	–0.001	0.000			
		T_0 (β_3)	0.041	0.009			
	Cluster 2 (N = 33)	Constant (β_0)	9,838.197	2,324.765	0.000	0.72	0.348
		C_{-1} (β_1)	0.548	0.217			
		T_{-1} (β_2)	–204.274	178.492			
		R_0 (β_3)	–41.212	14.105			
	Cluster 3 (N = 48)	Constant (β_0)	4,977.236	1,203.076	0.000	0.60	0.188
		C_{-1} (β_1)	0.400	0.110			
		R_0 (β_2)	–10.690	3.601			
		T_0 (β_3)	179.221	46.468			

C_{-1} – average consumption in the previous month (l/(customer.month)); T_{-1} – average temperature in the previous month (°C); T_0 – average temperature of the current month (°C); R_0 – rainfall of the current month (mm).

^aMLR model with logarithm transformation of the dependent variable (average monthly consumption).

conditions. As expected, the average monthly consumption increases with the temperature and decreases with the rainfall in all clusters, although with different ratios: Cluster 1 with $\hat{\beta}_2 = -0.001$ for rainfall and $\hat{\beta}_3 = 0.041$ for temperature; Cluster 2 with $\hat{\beta}_3 = -41.212$ for rainfall and $\hat{\beta}_4 = 353.398$ for temperature; and Cluster 3 with $\hat{\beta}_2 = -10.69$ for rainfall and $\hat{\beta}_3 = 179.221$ for temperature.

An example of the four plots used in MLR model testing is shown in Figure 10 for Cluster 3. The normal quantile-quantile plot of residuals (QQ plot) in Figure 10(a) shows that the regression errors are nearly normally distributed, since the majority of points are close to the line; however, there are also some candidates for outliers. Additionally, the residuals versus fitted value plot (Figure 10(b)) confirm the assumption of homoscedasticity; the errors seem to have constant variance, with the residuals scattered randomly around zero.

Figure 10(c) shows the square root of the standardized residuals (residuals rescaled to have a mean of zero and a variance of one) versus the fitted value. This plot is useful to see whether the variance is constant. From this plot, it does appear that the variance of the response slightly increases with the mean, but this effect appears to be null, if the outlier point is removed. To identify observations with potentially high influence, the plot of standardized residuals against leverage is shown in Figure 10(d). It is clear that there are at least two candidates for outliers (elements number 16 and 24).

MLR models have been validated using new DMA for the three clusters. The chosen DMA are from the same network areas, however corresponding to different years from those used to obtain the models. A relative error given by the absolute difference between the real and the estimated value divided by the real value was calculated. A DMA

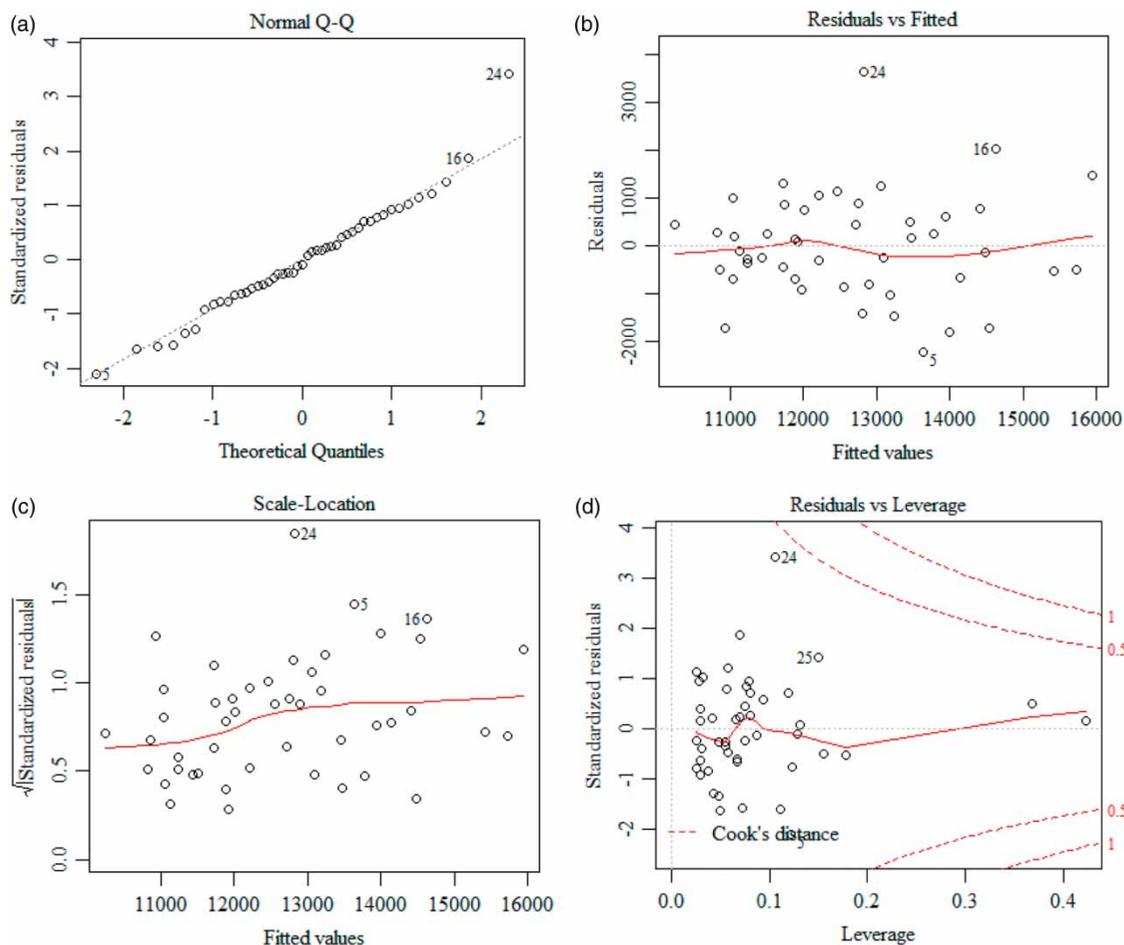


Figure 10 | MLR model testing for Cluster 3: (a) normal QQ plot, (b) residuals against fitted value, (c) standardized residuals against fitted value, and (d) standardized residuals against leverage points (Cook's distance).

Table 5 | Absolute relative errors for model validation

Clusters	DMA	Month	Year	Relative error (%)
Cluster 1	NO_Lis	December	2013	2.5
Cluster 2	SDM_Set	July	2014	2.6
	VEN_Set	October	2014	0.7
Cluster 3	FC_Lis	March	2013	7.3
	CS_Lis	May	2013	0.7

from Cluster 1 has been used to validate the model for December 2013. In Clusters 2 and 3, two DMA of each cluster were selected to validate the respective models for different months and years. Table 5 presents the relative errors, showing that the models have a good prediction performance with a relative error below 8%.

CONCLUSIONS

The current research is aimed at studying the influence of weather conditions on water consumption of DMA located in different Portuguese WDS. This allows improving the knowledge of urban water consumption, namely understanding the network sectors in which the urban consumption (domestic and non-domestic) is more sensitive to climate conditions. This aspect is important to improve the operational management of the WDS, given the predicted consumption scenarios for extreme weather events, namely increasing temperature. Additionally, in order to improve water demand management in situations of scarcity and to be able to establish measures to promote the efficient use of water, it is necessary to better understand which water uses are more influenced by weather conditions. The main contributions of this research are: the improved methodology using CA to segment DMA; the application of this methodology using an approach at the DMA level and with high resolution flow time-series (15 minutes); and the development, testing and validation of novel demand models between water consumption and weather variables.

The methodology proposed is a three-step procedure. This methodology allows collecting flow series without outliers and with a regular time-step of 15 minutes, to group different DMA with similar consumption billed characteristics, and to obtain MLR model between the average monthly consumption for the three clusters and weather conditions (temperature and

rainfall), taking into consideration the serial correlation of error terms, the autocorrelation and cross-correlation analysis between time-series and the logarithm transformation of the models.

The study of the influence of weather conditions on water consumption was carried out with 10 DMA in the south region of Portugal, selected from a set of 44 DMA. Results have shown that weather conditions, mainly temperature, have a significant influence on public billed consumption and domestic billed consumption in the third and fourth tariff levels. These contribute to the increase in water consumption, mainly the outdoor uses. Three clusters have been analyzed, and demand models for each cluster were developed. Demand models were validated, showing that the models have a good prediction performance with a relative error below 8%. Although the demand models are case-specific, they can be applied to DMA with billed consumption and socio-demographic characteristics similar to the ones analyzed herein.

ACKNOWLEDGEMENTS

The authors would like to thank the Portuguese water utilities that have provided consumption and infrastructure data for their systems.

REFERENCES

- Almeida, M. C., Vieira, P. & Ribeiro, R. 2006 Efficient Use of Water in Urban Sector (in Portuguese). IRAR, INAG, LNEC, Lisbon, Portugal.
- Arbués, F., Barberán, R. & Villanúa, I. 2004 Price impact on urban residential water demand: a dynamic panel data approach. *Water Resour. Res.* **40** (11), W11402.
- Babel, M., Gupta, A. D. & Pradhan, P. 2007 A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. *Water Resour. Manage.* **21** (3), 573–589.
- Beal, C. & Stewart, R. A. 2011 *South East Queensland Residential End Use Study: Final Report*. Urban Water Security Research Alliance, Brisbane, Australia.
- Behboudian, S., Tabesh, M., Falahnezhad, M. & Ghavanini, F. A. 2014 A long-term prediction of domestic water demand using preprocessing in artificial neural network. *J. Water Supply Res. Technol. Aqua* **63** (1), 31–42.

- Cabral, M. 2014 *Water demand projection in water distribution systems using a novel scenario planning approach*. MSc Thesis, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.
- Cole, M. A. 2004 *Economic growth and water use*. *Appl. Econ. Lett.* **11** (1), 1–4.
- Corbella, H. M. & Pujol, D. S. 2009 What lies behind domestic water use? A review essay on the drivers of domestic water consumption. *B. Asoc. Geogr. Esp.* **50**, 297–314.
- Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Roberson, J. A. 2012 *Urban water demand forecasting: a review of methods and models*. *J. Water Resour. Plann. Manage.* **140** (2), 146–159.
- Durbin, J. & Watson, G. S. 1950 *Testing for serial correlation in least squares regression: I*. *Biometrika* **37** (3/4), 409–428.
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R. & Mearns, L. O. 2000 *Climate extremes: observations, modeling, and impacts*. *Science* **289** (5487), 2068–2074.
- Farley, M. & Trow, S. 2003 *Losses in Water Distribution Networks: A Practitioner's Guide to Assessment, Monitoring and Control*. IWA Publishing, London.
- González, F. C. & Carranza, J. C. I. 2003 *Supply Manual of Canal de Isabel II*. Canal de Isabel II, Madrid.
- Gössling, S., Peeters, P., Hall, C. M., Ceron, J. P., Dubois, G. & Scott, D. 2012 *Tourism and water use: supply, demand, and security. An international review*. *Tourism Manage.* **33** (1), 1–15.
- Haasnoot, M., Middelkoop, H., Van Beek, E. & Van Deursen, W. P. A. 2011 *A method to develop sustainable water management strategies for an uncertain future*. *Sustain. Dev.* **19** (6), 369–381.
- Haugh, L. D. 2010 *Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach*. *J. Am. Stat. Assoc.* **71** (354), 378–385.
- Hoffmann, M., Worthington, A. & Higgs, H. 2006 *Urban water demand with fixed volumetric charging in a large municipality: the case of Brisbane, Australia*. *Aus. J. Agric. Resour. Econ.* **50** (3), 347–359.
- Höglund, L. 1999 *Household demand for water in Sweden with implications of a potential tax on water use*. *Water Resour. Res.* **35** (12), 3853–3863.
- House-Peters, L. A. & Chang, H. 2011 *Urban water demand modeling: review of concepts, methods, and organizing principles*. *Water Resour. Res.* **47** (5), W05401.
- INE 2012 *Census 2011: Final Results – Portugal*. National Statistics Institute, Lisbon, Portugal.
- IRAR 2009 *Tariff Formation for End-users of Drinking Water Supply, Urban Wastewater and Municipal Waste Management Services*. ERSAR, Lisbon, Portugal.
- Lee, J. & Chae, S. K. 2015 *Hourly water demand forecasting for micro water grids*. *J. Water Supply Res. Technol. Aqua* **65** (1), 12–17.
- Loh, M. & Coghlan, P. 2003 *Domestic Water Use Study*. Water Corporation, Perth, Western Australia.
- Loureiro, D. 2010 *Consumption Analysis Methodologies for the Efficient Management of Water Distribution Systems* (in Portuguese). PhD Thesis, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal.
- Loureiro, D., Amado, C., Martins, A., Vitorino, D., Mamade, A. & Coelho, S. T. 2015 *Water distribution systems flow monitoring and anomalous event detection: a practical approach*. *Urban Water J.* **13** (3), 242–252.
- Martinez-Espineira, R. 2002 *Residential water demand in the northwest of Spain*. *Environ. Resour. Econ.* **21** (2), 161–187.
- Miaou, S. P. 1990 *A class of time series urban water demand models with nonlinear climatic effects*. *Water Resour. Res.* **26** (2), 169–178.
- O'Brien, R. M. 2007 *A caution regarding rules of thumb for variance inflation factors*. *Qual. Quant.* **41** (5), 673–690.
- OECD 1999 *Household Water Pricing in OECD Countries*. OECD, Paris.
- Parandvash, G. H. & Chang, H. 2016 *Analysis of long-term climate change on per capita water demand in urban versus suburban areas in the Portland metropolitan area, USA*. *J. Hydrol.* **538**, 574–586.
- Pullinger, M., Anderson, B., Browne, A. L. & Medd, W. 2013 *New directions in understanding household water demand: a practices perspective*. *J. Water Supply Res. Technol. Aqua* **62** (8), 496–506.
- Rebelo, M., Loureiro, D., Santos, D., Coelho, S. T., Alegre, H. & Machado, P. 2008 *Characterization of Network Sectors in Water Distribution Systems: The Contribution of Social Sciences Supported by a Geographic Information System* (in Portuguese). In: *13th National Meeting of Basic Sanitation, Covilhã, Portugal*.
- Robinson, C. & Schumacker, R. E. 2009 *Interaction effects: centering, variance inflation factor, and interpretation issues*. *Multi. Linear Regress. Viewpoints* **35** (1), 6–11.
- Scargle, J. D. 1989 *Studies in astronomical time series analysis. III—Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data*. *Astrophys. J.* **343**, 874–887.
- Silva, J. F., Haie, N. & Vieira, J. 1995 *Analysis, Modeling in Cascading and Projection of Water Consumption*. Universidade do Minho, Lisbon, Portugal.
- SNIRH 2015 *Portuguese National Information System of Water Resources*. Available from: <http://snirh.pt/index.php?idMain=1&idItem=1.1> (accessed 16 December 2015).
- Vitorino, D., Loureiro, D., Alegre, H., Coelho, S. T. & Mamade, A. 2014 *In defense of the demand pattern, a software approach*. *Proc. Eng.* **89**, 982–989.
- Ward, J. H. 1963 *Hierarchical groupings to optimize an objective function*. *J. Am. Stat. Assoc.* **58**, 236–244.
- Weather Underground 2015 *Weather History and Data Archive*. Available from: www.wunderground.com/history/ (accessed 16 December 2015).