

# Demand scenario planning approach using regression techniques and application to network sectors in Portugal

Marta Cabral<sup>a,\*</sup>, Dália Loureiro<sup>b</sup>, Conceição Amado<sup>c</sup>,  
Aisha Mamade<sup>a,b</sup> and Dília Covas<sup>a</sup>

<sup>a</sup>*CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal*

*\*Corresponding author. E-mail: marta.f.cabral@tecnico.ulisboa.pt*

<sup>b</sup>*Urban Water Unit, National Civil Engineering Laboratory, Lisbon, Portugal*

<sup>c</sup>*CEMAT, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal*

---

## Abstract

This paper aims to develop and apply an approach for short-, medium- and long-term scenario planning. Construction of water demand scenarios is of the utmost importance for planning, design and operation of distribution systems, providing useful information to promote more efficient use of water resources. The proposed approach offers a new perspective for water utilities to obtain coherent and plausible descriptive scenarios as well as realistic water demand projections. This approach was applied to network sectors in existing water distributions systems, using extensive network flow data. Descriptive scenarios for water demand, considering trends and the uncertainties of the future, were obtained, together with improved Multiple Linear Regression (MLR) models, with the inclusion of the key variable temperature and by using historical billed consumption, socio-demographic and infrastructure data. In addition, water demand projections for short- and long-term periods were obtained through the inclusion of future trends in the MLR models. Results from MLR models show enhanced empirical relations between water demand variables and water consumption key factors. This approach was validated using different distribution systems and the results have demonstrated the efficiency of the developed models for the projection of water demand.

*Keywords:* Demand projections; Multiple Linear Regression models; Scenario planning; Water demand scenarios; Water distribution systems

---

## Introduction

Today, the world is witnessing global changes that may affect the provision of services from water utilities. The challenge for water utilities is the satisfaction of consumer demands, which presupposes their providing a sufficient quantity of water at a reasonable pressure and with an adequate quality

doi: 10.2166/wp.2019.029

© IWA Publishing 2019

(ISO, 2007). Scenario planning is a systemic approach and the central idea is to consider a variety of possible futures that include important uncertainties, rather than focussing on the accurate prediction of a single outcome (Peterson *et al.*, 2003). Usually, a *prediction* is a single interpretation of an event based on current best information, extrapolated to the future, considering that the future is relative similar to the past; a *scenario* represents a different version of an unknown future of an event (WBCSD, 2006). Scenario planning in water distribution systems (WDSs) allows forecasting the evolution of urban demand, availability or quality. Furthermore, it helps in the definition of the best design or rehabilitation alternative for a distribution system and in the establishment of appropriate management policies to ensure the continuity of distribution service with the minimum possible cost and risk (Silva *et al.*, 1995).

Urban water consumption includes domestic, industrial and service use (trade), as well as other uses (e.g., public consumption). The relevance and high weight of domestic consumption (indoor and outdoor uses) makes this component the most studied. Domestic water consumption may vary according to economic, socio-demographic, cultural or even religious variables, educational levels and responsiveness to conservation campaigns, consumers' physical capital, the predominant territorial urban form and climate variables (Bharati *et al.*, 2009; Corbella & Pujol, 2009; Qin *et al.*, 2017).

The development of scenarios for demand estimation should take into consideration key variables that influence water consumption, as well as relevant trends and uncertainties. According to Schoemaker (1995), trends are actual forces that have high impact in the field and which people believe to be highly predictable, such as well-known facts (e.g., an ageing population), while uncertainties are forces with a high impact in the field and with low predictability, such as unknown facts or solutions (e.g., a cure for cancer). Temporal and spatial scales are essential aspects to consider in the building of water demand scenarios. Alegre & Covas (2010) suggest that activities inside water utilities are organised in different decision levels: strategic, tactical and operational. At the strategic level, long-term objectives (10–20 years) are established for the entire system/organisation. At the tactical level, the objectives, the evaluation criteria and the metrics developed for a medium-time horizon (3–5 years) and a limited geographical scope are defined. The operational level establishes short-term actions (1–2 years).

During the last few decades, a considerable effort has been made to improve water demand estimation, to understand the different factors that influence consumption and to reduce the uncertainty of prediction models (Froukh, 2001). Previous studies on water resource management have demonstrated that scenarios are also useful to account for uncertainties associated with several key variables that affect the performance of water resource systems, including their effects on future water availability, water demand and water management strategies (Dong *et al.*, 2013). However, the demand estimation developed is known to be particularly challenging, due to the nature and the quality of available data, the existence of numerous key variables that affect water consumption and the multiplicity of temporal and spatial horizons involved (Donkor *et al.*, 2012). Most studies focus only on predicting water demand considering a short-term view (hours or days) based on the analysis of historical data (see, for example, Jain *et al.*, 2001; Pulido-Calvo *et al.*, 2003). On the other hand, long-term scenarios, using only qualitative data, focus on the study of water availability, rather than demand, considering a temporal scale longer than 60 years (see, for example, Chenoweth & Wehrmeyer, 2006).

The estimation of water demand can be used for the establishment of sustainable tariffs for water services, though several other factors besides water demand should also be considered for the establishment of tariffs, such as environmental and social concerns, governance and financial sustainability (Pinto &

Marques, 2016). In Portugal, a number of studies have been developed in this area (see, for example, Monteiro & Palma, 2011; Martins *et al.*, 2013). However, the influence of water demand on the establishment of tariffs will not be considered in this study.

This paper presents a scenario planning approach based on previous studies and combines two different methodologies: scenario planning, usually used to obtain descriptive scenarios for the long term, and Multiple Linear Regression (MLR) models, used to predict water demand based on the analysis of historical data. This approach considers a temporal scale that can range from 1 to 20 years combined with a spatial scale varying from District Metered Areas (DMAs) to entire WDSs. In this study, the proposed approach was applied to a set of DMAs distributed throughout Portugal with continuous network flow measurements allowing the following outcomes to be obtained:

1. descriptive scenarios for water demand, considering the trends and uncertainties of the future;
2. improved MLR models, with the inclusion of a short-term key variable (i.e., temperature) using historical billed consumption, socio-demographic and infrastructure data;
3. water demand projections for short and long terms, through the inclusion of future trends considered in the descriptive scenarios of the MLR models.

The paper is organised as follows. A scenario planning approach using regression techniques is first described, then a description of the case studies and the results obtained in the different modules are presented. The paper closes with a discussion of results and final considerations.

## Scenario planning approach

The scenario planning approach presented is a five-module procedure as shown in Figure 1, based on four existing scenario planning approaches (Von Reibnitz, 1988; Schwartz, 1991; Schoemaker, 1995; Peterson *et al.*, 2003). These four approaches are mostly used for qualitative data in order to obtain descriptive scenarios and are generally applied in organisational management to support decision making. They share common steps in scenario planning, such as the definition of scope, the

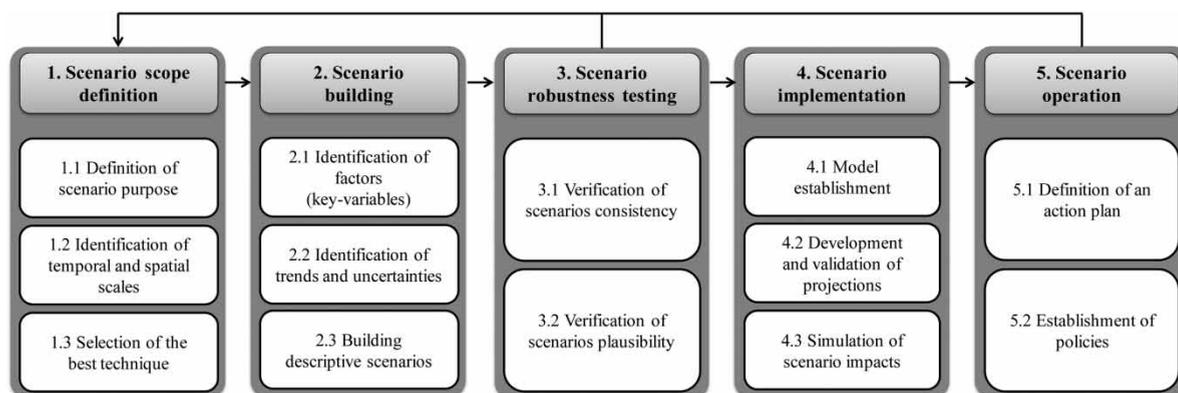


Fig. 1. Proposed approach for scenario planning in WDSs based on Von Reibnitz (1988), Schwartz (1991), Schoemaker (1995) and Peterson *et al.* (2003).

identification of key variables, trends and uncertainties and the building of descriptive scenarios. However, each approach contributes with specific steps. Von Reibnitz (1988) describes relevant steps such as consequence analysis of scenarios, development of core strategy and monitoring systems; Schwartz (1991) suggests that one of the steps in the approach is the study of the implications of built scenarios, whilst Schoemaker (1995) highlights the importance of checking the consistency and plausibility of built scenarios and of developing quantitative methods; Peterson et al. (2003) indicate that policy screening should be the last step in the scenario planning approach.

The first module (Scenario scope definition) is composed of three steps: 1.1 Definition of scenario purpose; 1.2 Identification of temporal and spatial scales; and 1.3 Selection of the best technique for scenario building. The first step focuses on the definition of the object of study (e.g., water demand, water availability, or both), the demand category (e.g., domestic consumption, public consumption, large consumers or other types) and expected projections (e.g., demand variables or demand patterns). Demand variables (e.g., peaking factors, minimum night demand or average daily demand) are organised into network sectors taking into consideration the infrastructure, as well as socio-demographic and billing factors. The second step aims to identify the decision level (i.e., strategic, tactical or operational) and the respective temporal and spatial scales. The third step focuses on the selection of the best technique for development of demand scenarios. Several techniques for scenario building in a broad context can be used. In the last decade, different schools of scenario techniques have emerged, such as ‘*La prospective*’ or normative/deterministic models, intuitive logics or social constructivist models, and probabilistic modified trend or rational/objectivist models (Wilkinson & Eidinow, 2008). In this study, the technique used to build the demand scenarios was developed by Godet (2006) from the *La prospective* school, also known as morphological analysis, and the objective is to present a different matrix for each temporal and spatial scale, using a mathematical and computer-based probabilistic approach.

The second module (Scenario building) includes three steps: 2.1 Identification of factors (key variables) that influence domestic water consumption; 2.2 Identification of trends and uncertainties associated with each key variable; and 2.3 Building descriptive scenarios. The trends and uncertainties associated with each key variable can result from existing scenario studies or univariate or multivariate linear regression analysis of historical data from the explanatory variables. After the identification of factors, trends and uncertainties, descriptive scenarios are built for different temporal and spatial scales identified in Module 1.

The third module (Scenario robustness testing) focuses on testing the robustness of the initially built scenarios, regarding consistency and plausibility. Consistency is the combination of logic in a scenario, ensuring that there is no built-in internal inconsistency that would undermine its credibility (Ratcliffe, 2000). Three types of consistency should be checked: trend, outcome and stakeholder consistency (Schoemaker, 1995). Plausibility implies that selected scenarios have to be capable of happening (Ratcliffe, 2000). Combinations that are not credible or possible should be eliminated and new scenarios should be developed until comprehensive internal consistency is achieved by a wide range of outcomes.

The fourth module (Scenario implementation) includes: 4.1 Model establishment; 4.2 Development and validation of projections; and 4.3 Simulation of scenario impacts. This last step identifies and quantifies the consequences of the different scenarios; however, it is not carried out in this study. The first step (Model establishment) is based on the methodology used to characterise consumption proposed by Loureiro et al. (2014, 2016), presented in Figure 2. The first step consists of data processing, data analysis and construction of MLR models. Different types of forecasting methods can be used to establish

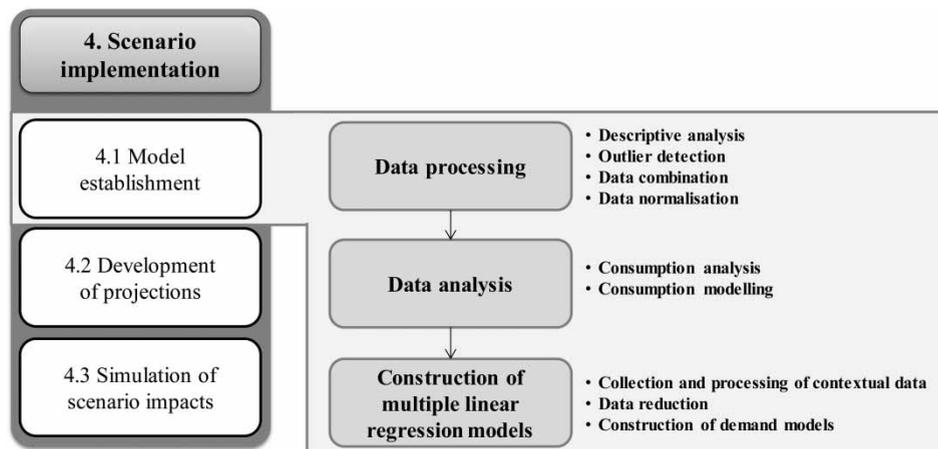


Fig. 2. Proposed steps for scenario implementation based on Loureiro *et al.* (2014, 2016).

water demand models, such as regression models, moving average and generalised autoregressive conditional heteroskedasticity (see, for example, Billings & Jones, 2008; Caiado, 2009). In this study, MLR analysis is used to establish the water demand models allowing empirical relations between consumption variables and the explanatory variables to be obtained. This analysis can imply autocorrelation and cross-correlation of the data as well as correlation of error terms. To avoid these implications, the obtained models were evaluated using the Durbin–Watson test and the global validation of linear model assumptions.

Network flow data processing includes descriptive analysis, outlier detection, data combination and data normalisation. Descriptive analysis aims to characterise the data and to identify potential outliers or less reliable data. The methodology developed by Loureiro *et al.* (2015) was used for outlier detection to identify and to remove outliers (outliers being observations that do not follow the statistical distribution of the data). Data combination includes removing large consumers and minimum night consumption (usually associated with water losses). Data normalisation consists of transforming the flow series into a regular time step (e.g., 15 minutes).

Network flow data analysis includes the analysis and modelling of consumption. Consumption analysis allows the consumption variables to be calculated, whilst consumption modelling allows the consumption patterns to be calculated, to evaluate consumption throughout the day.

The construction of MLR models includes the collection and processing of contextual data, the reduction of data and construction of demand models. Contextual data can be categorised into: clients and billed consumption; infrastructure; hydraulic operation; socio-demographic; consumption habits; facility characteristics; and climate. Collected and processed contextual data needs to be minimised for further analysis. Data reduction can be undertaken by two different processes: Cluster Analysis (CA) and Principal Component Analysis (PCA). MLR analysis is one of the methods most used for the construction of demand models using quantitative variables. The incorporation of identified trends represents a novel element of the proposed approach that allows water demand projections to be obtained and, finally, for the impacts of the different scenarios built into the network to be identified.

PCA and MLR analysis are used here to identify the most relevant variables and to establish the models. Regression analysis aims to develop empirical relations for consumption variables from the

principal components (PCs)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

where  $y$  are the dependent variables (e.g., consumption);  $x_1, x_2, \dots, x_p$  are the PCs or the independent variables (e.g., elderly families or public billed consumption);  $\beta_0, \beta_1, \dots, \beta_p$  are the regression coefficients, in which  $\beta_0$  is the constant term and  $\beta_1, \dots, \beta_p$  are coefficients that represent increases (positive value) or decreases (negative value) of dependent variables associated with a unit variation of the PCs; and  $\varepsilon$  is the random component that represents the disturbance or error term. Measures of goodness-of-fit calculated for the regression are (i) adjusted r-square ( $r_a^2$ ) and (ii)  $p$ -value, where  $r_a^2$  is a modification of r-square adjusted to the number of explanatory variables in a model, representing 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.

A Variance Inflation Factor (VIF) is used to measure the degree of multi-collinearity between independent variables in a regression model. Generally, a VIF should not exceed a value of 10 (Robinson & Schumacker, 2009). This factor is also used in MLR model construction. A study of the serial correlation of error terms was carried out using the Durbin–Watson test and global validation of linear model assumptions. The Durbin–Watson test computes residual autocorrelations, allowing conclusions to be drawn about the independence of error terms (Durbin & Watson, 1950). The global validation of linear model assumptions included global stat, skewness, kurtosis, link function and heteroscedasticity tests. All these tests show bootstrapped  $p$ -values, in which the null hypothesis is not rejected when the  $p$ -value is higher than 0.05.

The last module (Scenario operation) focuses on the definition of an action plan and the establishment of water management policies. A successful scenario planning effort should enhance the ability of people to cope with and take advantage of future changes, allowing the identification of more vulnerable policies or, conversely, being more robust to the uncertainties of the future. Decisions can be made, policies changed and management plans implemented to steer the system towards a more desirable future (Peterson et al., 2003). The definition of an action plan and the establishment of policies may foster the need for new scenarios to be implemented.

## Case studies

This analysis included two different case studies used for different purposes: (i) Case study 1 was used to apply the proposed scenario planning approach allowing descriptive scenarios, the MLR models and the respective projections to be obtained; and (ii) Case study 2 was used to validate the MLR models through a comparison of the different projections' values.

Case Study 1 consisted of network flow data from DMAs with domestic billed consumption higher than 50% of the total billed consumption. These DMAs correspond to a set of Portuguese WDSs distributed throughout the country, belonging to the following districts: Braga and Oporto (in the north region) and Lisbon and Setúbal (in the south region); see Figure 3. The total number of DMAs studied was 21 (twelve from the north region and nine from the south region) with flow data series ranging from the years 2010 to 2012 (three DMAs from 2010, twelve from the 2011 and six DMAs from the 2012).

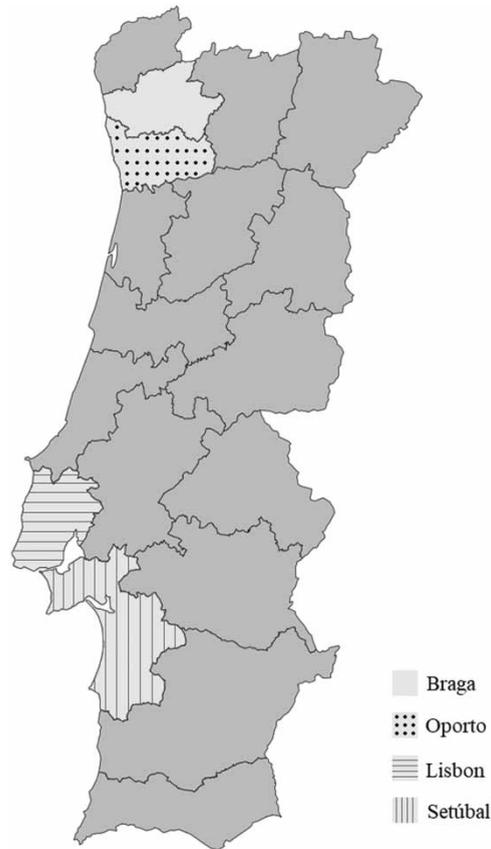


Fig. 3. Location of the Portuguese WDSs used in Case Study 1.

The general characteristics of analysed DMAs (including minimum and maximum values) are presented in Table 1.

The analysed DMAs have different characteristics regarding socio-demography, buildings, economic and individual mobility, and climate. The districts of Oporto and Braga are mainly characterised by rural areas with a lower density of population, while the districts of Lisbon and Setúbal present higher densities of population, corresponding mainly to urban areas. DMAs in the north present a higher percentage of buildings with one or two floors (88%) than DMAs from the south (51%), contributing to a lower number of service connections and network length in the DMAs belonging to the district of

Table 1. General characteristics, with minimum and maximum values, of analysed DMAs for the two case studies.

Characteristic and scope	Case Study 1 ( $N = 21$ )	Case Study 2 ( $N = 3$ )
Number of clients	441–5,185	826–1,176
Number of domestic clients	440–4,514	753–1,064
Number of inhabitants	1,940–15,898	4,820–5,147
Number of service connections	168–5,863	826–1,176
Network length [km]	6–152	26–88

Lisbon and Setúbal (Figure 4). Regarding family size, there are only 33% of families with one or two members in DMAs from the north, against 59% from the south.

Concerning economic and individual mobility, 42% of the workers living in DMAs in the north region are employed in the tertiary sector and only 7% of residents have a post-secondary school course or university degree, against 81% of workers and 16% who are graduates in DMAs in the south region. These higher percentages of employees in the tertiary sector and of graduates may lead to fewer conservative attitudes concerning the efficient use of water: Beal & Stewart (2011) suggest that employees in the tertiary sector and graduates have a higher water consumption in the summer due to outdoor uses, such as swimming pools and the irrigation of green spaces. Outdoor uses were also correlated to climate, mainly with temperature and precipitation. The north region is characterised by lower average annual temperature ( $T = 15^{\circ}\text{C}$ ) and higher average annual precipitation ( $P = 139\text{ mm}$ ) than the south region (where  $T = 19^{\circ}\text{C}$  and  $P = 58\text{ mm}$ ).

Case Study 2 was composed of three DMAs belonging to the district of Braga (north region of the country) with flow data series from 2012. Each DMA was identified with a code made up from abbreviations of the DMA name and district name. For instance, FAI\_Bra refers to a DMA in Braga district. The other DMAs studied were MAR\_Bra and PER\_Bra. As the characteristics of the DMAs used to validate the proposed approach needed to be within the range of DMAs used to apply the scenario planning approach (Case Study 1), only these three DMAs could be considered for validation.

## Methods and results

### Building descriptive scenarios

A large number of key variables were identified, collected and organised in categories: Economic, Socio-demographic, Climate, Infrastructure, Technological, and Regulations and ordinances (Mamade, 2013; Loureiro et al., 2016). These key variables were grouped by temporal scale (short-, medium- or

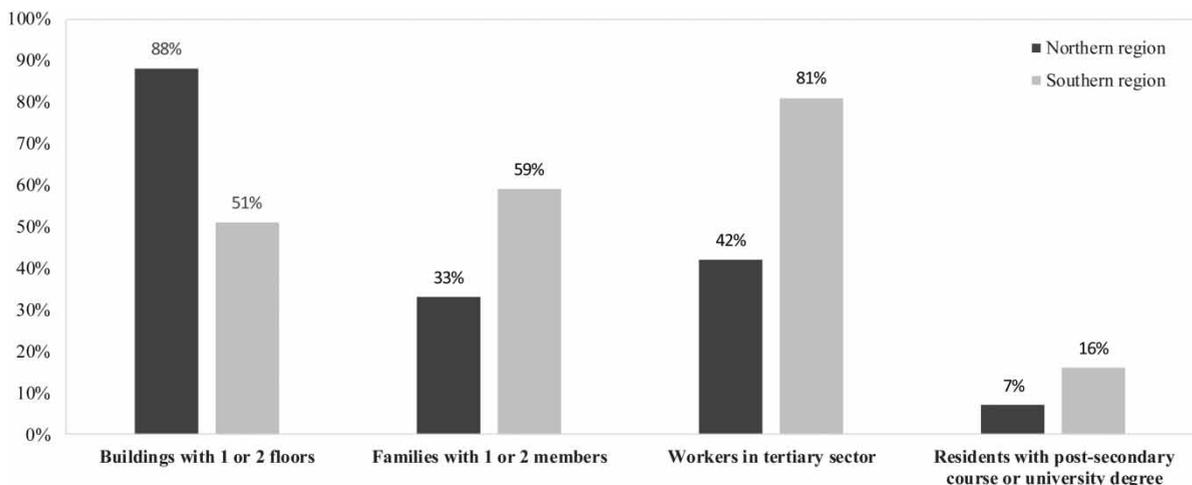


Fig. 4. Characterisation of Case Study 1 with regard to buildings, socio-demography, economic and individual mobility variables.

long-term) according to their influence on domestic water consumption in the region of Lisbon, Portugal. Short-term key variables were associated with seasonality (e.g., tourism, temperature and rainfall). Medium-term key variables were associated with the economy (e.g., water pricing), or system infrastructure, technologies, regulation and ordinances (e.g., water restrictions). Finally, long-term key variables were associated with socio-demographics (e.g., household size and housing typology).

Trends and uncertainties associated with key variables were collected and characterised for the south region of the country. Those came from scenario studies, such as the National Statistics Institute (INE) of Portugal and from the analysis of current and historical data from the explanatory variables. The scenario study from [Coelho et al. \(2008\)](#) forecasted an increase in the ageing population, including an increase of the elderly and of families with one or two members, and a decrease of families with three or four members in Portugal. Furthermore, historical data show an increase in inactive workers and university graduates.

Scenario building was illustrated with short-term (1 year) and long-term (20 year) scenarios. Short-term scenarios aimed to verify the influence of seasonality on demand through the temperature variable. Two different scenarios were built considering two different situations: (i) increase of temperature corresponding to an increase of public billed consumption and (ii) decrease of temperature corresponding to a decrease of public billed consumption (i.e., consumption associated with outdoor water uses in fountains, street washing, irrigation and cleaning of sewers). The uncertainty associated with scenarios corresponds to the values used to describe future trends. In these two scenarios, the maximum (Scenario 1) and minimum (Scenario 2) recorded values of average daily summer temperature and public billed consumption were used. These two scenarios project water demand using the monthly peaking factor and average demand per inhabitant.

The long-term scenarios aim to evaluate the influence of socio-demographic variables on demand. Three long-term scenarios were built. The first two scenarios showed the influence of elderly families on demand, using the trend values from [Coelho et al. \(2008\)](#). To exemplify the uncertainty, Scenario 4 assumed that trend values are only 10% of what is projected, while Scenario 5 used corrected projections. In the third scenario, the increase in temperature was also considered and, consequently, so was public billed consumption ([Figure 5](#)). These three scenarios were used to project water demand using minimum night demand per client.

A group of specialists in sanitary engineering analysed and discussed the scenarios built taking into consideration the main trends and uncertainties about the future in the south region of the country. Robustness testing was developed through brainstorming, allowing the coherence and plausibility of the developed scenarios to be assessed.

### *Construction of Multiple Linear Regression models*

Results from a previous work by [Mamade \(2013\)](#) about PCA were used, including the PCs and variables obtained. A temperature variable was also incorporated in the current study. Since significant regional differences were identified, PCA was separately carried out for the north and the south regions. Five PCs and five variables were considered in the regression analysis. The five PCs were ‘elderly families’ and ‘individual mobility’ (from the Socio-demographic category), ‘pipe material’ and ‘pipe size’ (from the Infrastructure category) and ‘domestic consumption’ (from the Billing category). The five variables were: ‘commercial and industry billed consumption’, ‘public billed consumption’, ‘collective billed consumption’, ‘average daily summer temperature’ and ‘region’. The values of the PCs are

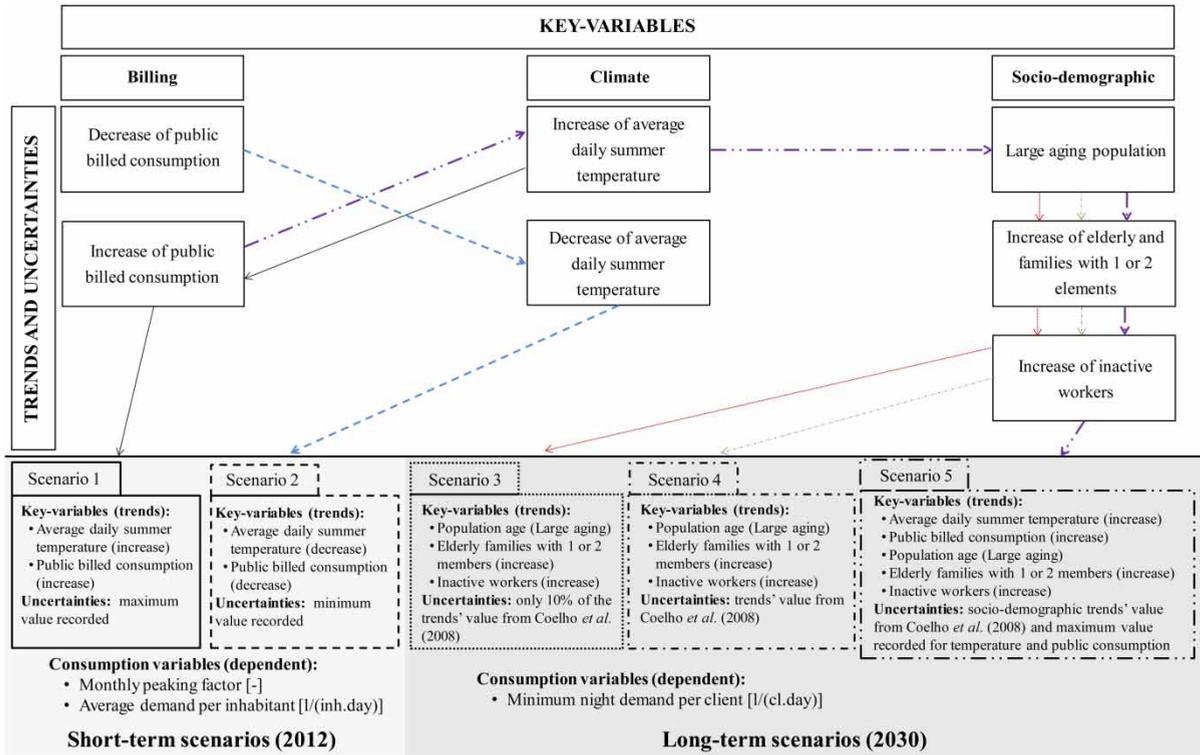


Fig. 5. Building descriptive scenarios for the short (Scenarios 1–2) and long term (Scenarios 3–5).

called factor scores, which can be interpreted geometrically as projections of observations of the PCs (Abdi & Williams, 2010). Thus, the values of the PCs may be positive or negative. The input values in regression models depend on the region aiming to implement the demand scenarios; nonetheless, the chosen values must be included in the range of their PCs or variable.

MLR analysis was used to improve the existing demand models by incorporation of the temperature variable. However, it was initially necessary to verify the correlation between the independent and dependent variables. This step was carried out for the south region (nine DMAs) and then for the north and south regions together (21 DMAs) in order to choose the independent (key) variables that best explain the water consumption for the different consumption variables. Correlation matrices were calculated as they show the correlation coefficients between the dependent and independent variables; a correlation coefficient higher than 0.5 (positive or negative) indicates that a given key variable is related to another dependent variable, and therefore selected.

A correlation matrix between the different independent variables was also computed in order to confirm the low correlation between them and their independence in the regression model. If a correlation coefficient is high, the variables are not considered independent, and one of them should be discarded. This analysis showed high correlation between two variables – domestic consumption and average daily summer temperature – with a correlation coefficient of  $-0.78$  and, in some regressions, the VIF was close to 10. Thus, these two variables have large multi-collinearity and should not be considered simultaneously as independent variables.

Descriptive statistics (including maximum, minimum, average, standard deviation, 1st quartile, 2nd quartile–Median and 3rd quartile) for the PCs and variables used for the construction of demand prediction models are presented in Table 2. Loureiro *et al.* (2016) explained how the PCs are constructed using the PCA technique and the meaning of the values (Table 3). As an example, for the pipe material the value can vary between  $-1.485$  and  $3.791$  (the negative values being associated with plastic pipes and positive values with asbestos cement pipes). Region is a dummy variable, which means that it can only present two values, zero or one, for the north or south region, respectively. Descriptive statistics were also obtained for the following dependent variables (Table 4):

- monthly peaking factor: the ratio between the maximum monthly consumption and the average annual consumption;
- average consumption per inhabitant: the ratio between the average consumption and the number of inhabitants – in litres per inhabitant and day [l/(inh.day)];
- median consumption per inhabitant: ratio between the median consumption and the number of inhabitants – in litres per inhabitant and day [l/(inh.day)];
- minimum night consumption per client: ratio between the minimum night consumption and the number of inhabitants – in litres per client and day [l/(cl.day)];
- night consumption per service connection: ratio between the minimum night consumption and the number of service connections – in litres per service connection and hour [l/(sc.h)].

The importance of incorporating the temperature variable was assessed through the comparison of results obtained by two demand models. The first had good results without adding the temperature. The second introduced the temperature variable but discarded the domestic consumption variable.

Since different behaviours for the north and the south region were observed, initially MLR analysis was separately carried out for the south region and, then, for both regions, introducing a new variable (the region). Better results were obtained when considering only the south region, since the temperature

Table 2. Descriptive statistics from PCA results for the principal components (PCs) and independent variables for the 21 DMAs.

Descriptive statistics	Principal components (PCs)				Variables				Average daily summer temperature, $T$ [°C]
	Elderly families, $E$ [-]	Individual mobility, $I$ [-]	Pipe material, $P_m$ [-]	Pipe size, $P_s$ [-]	Commercial and industry billed consumption, $C$ [%]	Public billed consumption, $P$ [%]	Collective billed consumption, $C_b$ [%]	Region, $R$ [-]	
Maximum	1.59	0.79	3.79	0.99	20.94	22.84	14.13	1	26.09
Minimum	2.22	-1.11	-1.49	-3.72	0.00	0.00	0.00	0	19.19
Average	0.01	-0.18	0.57	-0.32	6.78	4.13	1.98	-	22.37
Standard deviation	1.07	0.56	1.33	0.96	5.53	6.25	3.12	-	2.29
1st Quartile	-0.82	-0.59	-0.412	-0.53	1.27	0.00	0.00	-	20.62
2nd Quartile	0.09	-0.39	0.22	-0.18	6.84	0.00	1.21	-	22.43
3rd Quartile	0.95	0.24	1.2	0.23	9.13	7.84	2.75	-	24.40

Table 3. Initial independent variables considered for PCA in order to obtain the principal components (adapted from Loureiro et al., 2016).

Principal components	Initial independent variables
Elderly families	Inactive workers, elderly, families with adolescents, families with 1–2 members, families with 3–4 members
Individual mobility	Economic mobility, university graduates, people with 12 years of education
Pipe material	Plastic pipes, AC pipes, service connection density
Pipe size	% diameter 110–310, % diameter $\leq 110$

Table 4. Descriptive statistics for consumption variables (dependent variables) for the 21 DMAs.

Descriptive statistics	Monthly peaking factor [-]	Average consumption per inhabitant [l/(inh.day)]	Median consumption per inhabitant [l/(inh.day)]	Minimum night consumption per client [l/(cl.day)]	Night consumption per service connection [l/(sc.h)]
Maximum	1.85	290.12	305.29	707.12	65.48
Minimum	1.07	34.74	30.53	27.01	1.56
Average	1.26	126.03	126.20	214.88	15.67
Standard deviation	0.18	80.85	80.66	181.86	17.90
1st Quartile	1.12	60.59	59.49	85.44	2.47
2nd Quartile	1.22	99.84	98.71	138.34	5.66
3rd Quartile	1.34	205.30	206.97	321.62	23.28

Note: l/(inh.day) = litres per inhabitant and day; l/(cl.day) = litres per client and day; l/(sc.h) = litres per service connection and hour.

has a higher influence on the consumption in this region due to higher average temperatures and higher economic and individual mobility. The results of the MLR models are presented in Table 5, showing a significant improvement of the MLR models with the temperature variable, with regard to the adjusted r-squared and significance of the model ( $p$ -value). Average daily summer temperature is a more important explanatory variable of water demand than the domestic consumption variable. Results of the new MLR model in the south region show a good adjustment with  $\bar{r}_a^2 = 0.84$  (average coefficient of determination) and low  $p$ -value, increasing the significance of the model.

Consumption values increase with temperature: this is proven by the regression coefficients of the explaining component of temperature being positive and high (for example,  $\beta_4 = 56.191$  for the average consumption per inhabitant –  $\bar{Q}_{inh}$  – and  $\beta_5 = 81.972$  for the minimum night consumption per client –  $Q_{min}$ ). On the other hand, the  $\beta_0$  for the majority of the consumption variables in this study is negative; this regression coefficient represents the value of dependent variable when all the independent variables assume a null value. This result has no practical significance, since the values used for regression analysis of each independent variable do not incorporate the zero value. Besides, in some independent variables, the zero value, such as average daily summer temperature, cannot be introduced. Thus, there is no possibility that the value of the consumption variable can be equal to  $\beta_0$ .

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 this test is also presented in

Table 5. Results of the MLR analysis, with and without the temperature variable.

Region	Dependent variable	With temperature			Without temperature (Mamade, 2013)
		Linear functions	$r_a^2$ (p-value)	Durbin-Watson test p-value	$r_a^2$ (p-value)
South	$M_{pf}$ [-]	$M_{pf} = -0.474 - 0.048E - 0.001P + 0.061P_m + 0.071T$	0.68 (0.070)	0.216	0.58 (0.116)
	$\bar{Q}_{inh}$ [l/(inh.day)]	$\bar{Q}_{inh} = -1154.543 + 67.876I + 5.787C + 20.557P_m + 56.191T$	0.93 (0.004)	0.106	0.81 (0.027)
	$Q_{med\_inh}$ [l/(inh.day)]	$Q_{med\_inh} = -1012.830 + 66.705I + 7.002C + 24.947P_m + 49.393T$	0.94 (0.003)	0.148	0.86 (0.014)
	$Q_{min}$ [l/(cl.day)]	$Q_{min} = -2018.607 - 41.395E + 21.673C + 9.607P - 38.592P_s + 81.972T$	0.98 (0.003)	0.858	0.86 (0.038)
	$Q_{nig}$ [l/(sc.h)]	$Q_{nig} = -71.475 + 10.564E + 1.257C + 0.846P - 12.376P_s + 4.096T$	0.90 (0.017)	0.814	0.87 (0.036)
North and South	$M_{pf}$ [-]	$M_{pf} = -0.011 - 0.060I - 0.133R + 0.059T$	0.51 (0.002)	0.598	0.18 (0.098)

Note: **Dependent variables:**  $M_{pf}$  = Monthly peaking factor;  $\bar{Q}_{inh}$  = Average consumption per inhabitant;  $Q_{med\_inh}$  = Median consumption per inhabitant;  $Q_{min}$  = Minimum night consumption per client;  $Q_{nig}$  = Night consumption per service connection;  $M_{pf}$  = Monthly peaking factor. **Independent variables:**  $E$  = Elderly families;  $P$  = Public billed consumption;  $P_m$  = Pipe material;  $T$  = Average daily summer temperature;  $I$  = Individual mobility;  $C$  = Commercial and industry billed consumption;  $P_s$  = Pipe size;  $R$  = Region.

l/(inh.day) = litres per inhabitant and day; l/(cl.day) = litres per client and day; l/(sc.h) = litres per service connection and hour.

Table 5 and, for all models, this  $p$ -value is higher than 0.05, corresponding to the value at which the null hypothesis is not rejected. Also, in all consumption variables, the assumptions in global validation of multiple linear models were satisfied, including the assumptions of global stat, skewness, kurtosis, link function and heteroscedasticity.

### Development and validation of projections

Projection development consists of incorporating the trends of each independent variable into the MLR models for the south region, belonging to the districts of Lisbon and Setúbal. In the short-term scenario, the monthly peaking factor and the average demand per inhabitant were chosen as variables for projecting water demand for the following year, as presented in Table 6. The value of 1.22 in the monthly peaking factor, corresponding to the current state (2011, being the year analysed), was calculated by the average value of monthly peaking factors for the nine DMAs in the south region. These scenarios do not consider changes in elderly families and pipe material (using only the values of 2011), considering only the influence of average daily summer temperature and, consequently, of public billed consumption. The values of average daily summer temperature considered were 24.40°C and 22.43°C corresponding to the maximum and minimum values, respectively, for temperature in the south region. Results of the two constructed scenarios for the two demand variables show large variability, highlighting that small changes in the average daily summer temperature can cause significant changes in demand.

Long-term scenarios allow the minimum night demand per client to be projected to 2030. The study of this demand variable is essential for the assessment of water losses. These scenarios do not consider changes in commercial and industry billed consumption nor in pipe size (using values from 2011). First, two scenarios showed the influence of elderly families on the demand and the obtained projections presented lower values of minimum night demand per client than the current state, due to the increase in numbers of elderly people, small families (one or two members) and inactive workers. In the third scenario, the increase in average daily summer temperature and, consequently, of public billed consumption was also considered. The projected value of this scenario is far higher than the other two scenarios and

Table 6. Results of short- and long-term scenario construction for the south region.

Temporal scale	Dependent variable	Current state (2011)	Projection (Scenario)	Linear functions
Short-term scenarios (2012)	Monthly peaking factor [-]	1.22	1.30 (Scenario 1) 1.17 (Scenario 2)	$M_{pf} = -0.474 - 0.048E - 0.001P + 0.061P_m + 0.071T$
	Average demand per inhabitant [l/(inh.day)]	140.77	216.63 (Scenario 1) 105.94 (Scenario 2)	$\bar{Q}_{inh} = -1154.543 + 67.876I + 5.787C + 20.557P_m + 56.191T$
Long-term scenarios (2030)	Minimum night demand per client [l/(cl.day)]	175.46	148.66 (Scenario 3) 151.71 (Scenario 4) 269.06 (Scenario 5)	$Q_{min} = -2018.607 - 41.395E + 21.673C + 9.607P + 38.592P_s + 81.972T$

Note:  $E$  = Elderly families;  $P$  = Public billed consumption;  $P_m$  = Pipe material;  $T$  = Average daily summer temperature;  $I$  = Individual mobility;  $C$  = Commercial and industry billed consumption;  $P_s$  = Pipe size.  $l/(inh.day)$  = litres per inhabitant and day;  $l/(cl.day)$  = litres per client and day.

the current state, highlighting that minor changes in average daily summer temperature can cause significant changes in demand.

*Monthly peaking factor* was chosen as the demand variable to validate the proposed approach, using the three DMAs from Braga district (Case Study 2). The predictive interval of this variable was obtained using the software, R, and it can range from 1.108 to 1.559. Therefore, the monthly peaking factor of the three DMAs for 2012 must be in the range of the interval of the MLR analysis. The obtained linear function was calculated using the 21 DMAs, including DMAs from the north and south regions (Table 5). Monthly peaking factor can be expressed as:

$$M_{pf} = -0.011 - 0.060I - 0.133R + 0.059T \quad (2)$$

This function can be used to obtain the projected monthly peaking factor for 2012 and its projected value is independent of the DMAs' general characteristics, only varying with the following variables: individual mobility ( $I$ ), region ( $R$ ) and average daily summer temperature ( $T$ ). The result is a single projected value for the three analysed DMAs, obtained taking  $18.7^{\circ}\text{C}$  for the average daily summer temperature, corresponding to the district of Braga; average daily summer temperature can range from  $18.7^{\circ}\text{C}$  to  $24.4^{\circ}\text{C}$ , corresponding to the districts of Braga and Setúbal, respectively. For individual mobility,  $I = -0.382$  (this value being negative because it is obtained using PCA) and for region,  $R = 0$  (the value 0 is considered for the north region and the value 1 for the south region of the country) were taken to be the values for the analysed year of 2011. The projected monthly peaking factor for 2012 is 1.115 for the three analysed DMAs, this value being included in the predictive interval. The projected value is closer to the lower limit of the intervals, since it uses the lower value of average daily summer temperature (this assumption is due to the use of DMAs from the north region where the temperature is lower). The calculated monthly peaking factor from the flow series of the three DMAs are: FAI\_Bra = 1.1, MAR\_Bra = 1.3, PER\_Bra = 1.3. All these calculated values are included in the interval (Figure 6), showing the efficiency of MLR models constructed for the projection of water demand.

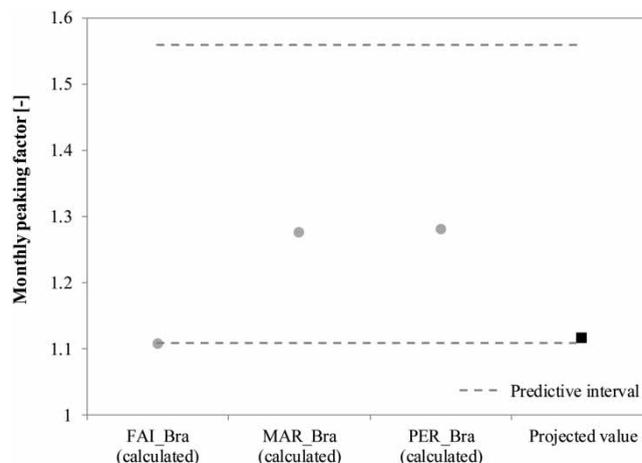


Fig. 6. Results of approach validation, with a predictive interval, monthly peaking factor values from the three analysed DMAs and projected value.

## Conclusions

This research aimed to develop and validate an approach to water demand scenario planning which considers different temporal and spatial scales. This approach allows water demand scenarios to be obtained for short- (up to 1 year), medium- (3–5 years) and long-term (10–20 years) periods in a distribution system or in smaller areas (e.g., network sectors or DMAs). This approach incorporated trends and uncertainties associated with key variables dependent on local conditions (e.g., socio-demographic, infrastructure, non-domestic billed consumption, temperature and region) in MLR models to obtain water demand projections.

During the application of the proposed approach, five descriptive scenarios were established, taking into consideration the identified factors that influence water consumption in Portugal and the trends and uncertainties in the south region of the country. Existing MLR models were improved using a Case Study of 21 predominantly domestic DMAs using the following key variables: ‘elderly families’, ‘individual mobility’, ‘pipe material’, ‘pipe size’, ‘public billed consumption’, ‘commercial and industry billed consumption’, ‘region’ and ‘average daily summer temperature’. The inclusion of temperature has significantly improved existing MLR models, proved by the obtained values of goodness-of-fit measures (i.e., adjusted r-squared and *p*-value). The MLR models obtained are applicable to DMAs in Portugal, which need to be within the range of the general characteristics of the DMAs used to apply the scenario planning approach.

Three different DMAs were used to validate the proposed approach using monthly peaking factor as the demand variable to be projected. The projected and calculated monthly peaking factors were included within the predictive interval, showing the efficiency of demand models constructed for the projection of water demand.

The main innovative contribution of this work is the development and validation of an approach to water demand scenario planning combining two different methodologies: scenario planning and MLR models. The combination of these methodologies allows more robust water demand projections to be obtained as a result of descriptive scenarios through the inclusion of trends and uncertainties in the MLR models. One of the main objectives of this approach is to build demand scenarios and not to indicate the most likely demand scenario for a given temporal and spatial scale, since all scenarios are plausible but with different occurrence probabilities. This approach allows knowledge of domestic water demand to be increased, which is of the utmost importance for the planning, design and operation of WDSs.

For future research, the authors recommend the application of the proposed approach in other countries, as well as an in-depth study of the role of water demand planning in the short- and long-term for water tariff sustainability.

## Acknowledgements

The authors would like to thank the Portuguese water utilities who provided consumption and infrastructure data from their systems, anonymously and confidentially.

## References

- Abdi, H. & Williams, L. J. (2010). *Principal component analysis*. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(4), 433–459.

- Alegre, H. & Covas, D. (2010). *Patrimonial Management of Infrastructure in Urban Water Systems: an Approach Centered on Rehabilitation (in Portuguese)*. ERSAR, LNEC, Lisbon, Portugal.
- Beal, C. & Stewart, R. A. (2011). *South East Queensland Residential End Use Study*. Final Report 47, Urban Water Security Research Alliance.
- Bharati, L., Smakhtin, V. U. & Anand, B. K. (2009). Modelling water supply and demand scenarios: the Godavari–Krishna inter-basin transfer, India. *Water Policy* 11(S1), 140–153.
- Billings, R. B. & Jones, C. V. (2008). *Forecasting Urban Water Demand*. American Water Works Association, Denver.
- Caiado, J. (2009). Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal of Hydrologic Engineering* 15(3), 215–222.
- Chenoweth, J. L. & Wehrmeyer, W. (2006). Scenario development for 2050 for the Israeli/Palestinian water sector. *Population and Environment* 27(3), 245–261. doi:10.1007/s11111-006-0021-6.
- Coelho, E., Magalhães, M., Peixoto, J. & Bravo, J. (2008). *Resident Population Projections in Portugal 2008–2060 (in Portuguese)*. Statistical National Institute (INE), Lisbon, Portugal.
- Corbella, H. M. & Pujol, D. S. (2009). What lies behind domestic water use? A review essay on the drivers of domestic water consumption. *Boletín De La Asociación De Geógrafos Españoles* 50, 297–314.
- Dong, C., Schoups, G. & van de Giesen, N. (2013). Scenario development for water resource planning and management: a review. *Technological Forecasting and Social Change* 80(4), 749–761.
- Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Roberson, J. A. (2012). Urban water demand forecasting: a review of methods and models. *Journal of Water Resources Planning and Management* 140(2), 146–159. doi: 10.1061/(ASCE)WR.1943-5452.0000314.
- Durbin, J. & Watson, G. S. (1950). Testing for serial correlation in least squares regression. II. *Biometrika* 37(3/4), 409–428.
- Froukh, M. L. (2001). Decision-support system for domestic water demand forecasting and management. *Water Resources Management* 15(6), 363–382. doi: 10.1023/A:1015527117823.
- Godet, M. (2006). *Creating Futures: Scenario Planning as A Strategic Management Tool*. Economica, Paris.
- International Organization for Standardization (ISO) (2007). *Activities Relating to Drinking Water and Wastewater Services – Guidelines for the Management of Drinking Water Utilities and for the Assessment of Drinking Water Services (ISO 24512: 2007)*. ISO, Geneva, Switzerland.
- Jain, A., Varshney, A. K. & Joshi, U. C. (2001). Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management* 15(5), 299–321.
- Loureiro, D., Alegre, H., Coelho, S., Martins, A. & Mamade, A. (2014). A new approach to improve water loss control using smart metering data. *Water Science & Technology: Water Supply* 14(4), 618–625.
- 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 Journal*, 1–11. doi: 10.1080/1573062X.2014.988733.
- Loureiro, D., Mamade, A., Cabral, M., Amado, C., Coelho, S. T., Alegre, A. & Covas, D. (2016). A comprehensive approach for spatial and temporal water demand profiling in network sectors. *Water Resources Management* 30(10), 3443–3457.
- Mamade, A. Z. (2013). *Profiling Consumption Patterns Using Extensive Measurements – A Spatial and Temporal Forecasting Approach for Water Distribution Systems*. MSc Thesis, Universidade de Lisboa, Lisbon, Portugal.
- Martins, R., Cruz, L., Barata, E. & Quintal, C. (2013). Assessing social concerns in water tariffs. *Water Policy* 15(2), 193–211.
- Monteiro, H. & Roseta-Palma, C. (2011). Pricing for scarcity? an efficiency analysis of increasing block tariffs. *Water Resources Research* 47(6), W06510. doi:10.1029/2010WR009200.
- Peterson, G. D., Cumming, G. S. & Carpenter, S. R. (2003). Scenario planning: a tool for conservation in an uncertain world. *Conservation Biology* 17(2), 358–366. doi: 10.1046/j.1523-1739.2003.01491.x.
- Pinto, F. S. & Marques, R. C. (2016). Tariff suitability framework for water supply services. *Water Resources Management* 30(6), 2037–2053.
- Pulido-Calvo, I., Roldán, J., López-Luque, R. & Gutiérrez-Estrada, J. (2003). Demand forecasting for irrigation water distribution systems. *Journal of Irrigation and Drainage Engineering* 129(6), 422–431.
- Qin, H., Cai, X. & Zheng, C. (2017). Water demand predictions for megacities: system dynamics modelling and implications. *Water Policy* 20(1), 53–76.
- Ratcliffe, J. (2000). Scenario building: a suitable method for strategic property planning? *Property Management* 18(2), 127–144.

- Robinson, C. & Schumacker, R. E. (2009). Interaction effects: centering, variance inflation factor, and interpretation issues. *Multiple Linear Regression Viewpoints* 35(1), 6–11.
- Schoemaker, P. J. (1995). Scenario planning: a tool for strategic thinking. *Sloan Management Review* 36(2), 25–40.
- Schwartz, P. (1991). *The Art of the Long View*. Doubleday Currency, New York.
- Silva, J. F., Haie, N. & Vieira, J. (1995). *Analysis, Modeling in Cascading and Projection of Water Consumption*. Universidade do Minho, Lisbon, Portugal.
- Von Reibnitz, U. H. (1988). *Scenario Techniques*. McGraw-Hill, New York.
- Wilkinson, A. & Eidinow, E. (2008). [Evolving practices in environmental scenarios: a new scenario typology](#). *Environmental Research Letters* 3(4), 045017.
- World Business Council for Sustainable Development (WBCSD) (2006). *Business in the World of Water: WBCSD Water Scenarios to 2025*. WBCSD, Conches-Geneva, Switzerland

Received 12 February 2018; accepted in revised form 7 November 2018. Available online 8 February 2019