JOURNAL OF WATER AND LAND DEVELOPMENT
e-ISSN 2083-4535

Polish Academy of Sciences (PAN), Committee on Agronomic Sciences
Institute of Technology and Life Sciences (ITP)

JOURNAL OF WATER AND LAND DEVELOPMENT
2021, No. 49 (IV–VI): 219–228. https://doi.org/10.24425/jwld.2021.137115.137115

Available (PDF): https://www.itp.edu.pl/JWLD; http://journals.pan.pl/jwld

Received 29.05.2020
Reviewed 16.08.2020
Accepted 20.10.2020

Assessing domestic factors determining water consumption in a semi-arid area (Sedrata City) using artificial neural networks and principal component analysis

Menal ZEROUAL[1)], Azzedine HANI[1)], Amir BOUSTILA[2)]

[1)] University of Badji Mokhtar, Faculty of Earth Sciences, Laboratory of water resource and sustainable development, BP 12 / 23000 Annaba, Algeria
[2)] University of Badji Mokhtar, Faculty of Earth Sciences, Laboratory of natural resource and development, Annaba, Algeria

For citation: Zeroual M., Hani A., Boustila A. 2021. Assessing domestic factors determining water consumption in a semi-arid area (Sedrata City) using artificial neural networks and principal component analysis. Journal of Water and Land Development. No. 49 (IV–VI) p. 219–228. DOI 10.24425/jwld.2021.137115.

Abstract

The growing demand for fresh water and its scarcity are the major problems encountered in semi-arid cities. Two different techniques have been used to assess the main determinants of domestic water in the Sedrata City, North-East Algeria: principal component analysis (PCA) and artificial neural networks (ANNs). To create the ANNs models based on the PCA, twelve explanatory variables are initially investigated, of which nine are socio-economic parameters and three physical characteristics of building units. Two optimum ANNs models have been selected where correlation coefficients equal to 0.99 in training, testing and validation phases. In addition, results demonstrate that the combination of socio-economic parameters with physical characteristics of building units enhances the assessment of household water consumption.

Key words: *artificial neural networks, domestic water use determinants, household water consumption, principal component analysis, semi-arid area*

INTRODUCTION

In this century, the supply of sufficient quantity of fresh water represents one of the major issues. In developed countries, several authors have extensively investigated water consumption patterns, whilst in developing countries no in-depth studies have been implemented [NAUGES, WHITTINGTON 2009]. According to UNESCO's 2018 report, with population increase, water demand has been growing by 20% per year.

Accordingly, the determination of parameters governing domestic water use is an important factor for designing water demand patterns, optimizing water distribution systems and facilitating future studies in urban planning. The understanding of water usage patterns helps to generate predictive models and consequently has a positive environmental impact. According to papers on the subject, researchers distinguish socio-economic parameters [FAN *et al*. 2017; HUSSEIN *et al.* 2016; MATOS *et al.* 2014], household habits, physical characteristics of buildings and climatic factors governing the use of water.

As regards socio-economic factors, FIELDING *et al.* [2012] demonstrated that water use is affected by household income and water price. Researchers, such as KENNEY *et al.* [2008] and GRAFTON *et al.* [2011], found that households with higher income use more water; However, BEAL *et al.* [2011] have shown the opposite. In addition to the family income, water use is related to a household size [ARBUÈS *et al.* 2010; GATO-TRINIDAD *et al.* 2011; MAKKI *et al.* 2015], and the number of women in a house [MU *et al.* 1990]. Furthermore, the age distribution of residents should also be considered in the analysis [MAKKI *et al*. 2011]. BALLING *et al.* [2008] have precisely demonstrated that children use more water than other age groups.

Other studies outlined that the number of adolescents in a household is the key variable of per capita water demand

[Aquacraft 2015]. Additionally, other studies reported that both older people [Beal *et al*. 2011] and retired individuals [Willis *et al*. 2009] also use more water. According to Kenney *et al.* [2008], the increase in water consumption is proportional to the combination of household income, wealth and age. On the contrary, Wentz *et al.* [2014] found that the age of people was not a significant factor affecting water consumption. Following from this, a higher education level is a driver for higher water use per capita [Makki *et al.* 2015]. Previous results show that the water consumption in households consisting of working people is higher than that in houses with retired residents [Makki *et al.* 2011].

Household water habits are a crucial parameter in water consumption forecasting. Recently, more attention has been paid to habits like showering, washing clothes, dishwashing and bathing [Kenney *et al.* 2008; Maki *et al.* 2015; Xue *et al.* 2017].

Factors like housing type (single or collective apartments), age of buildings, lot size (occupied surface) and garden size [House-Peters *et al.* 2011] are referred in literature as physical building characteristics and have been proven to influence water consumption. Generally, climatic variables like precipitation and temperature could affect water use as well [Arbués *et al.* 2010; Haque *et al.* 2015].

More recently, artificial intelligence techniques (AI) are increasingly often applied due to their ability to perform non-linear curve fitting and analysis to very complex data sets. AI is used where we encounter noisy data and the structure of the model remains unknown [Tiwari *et al*. 2013; Tiwari *et al.* 2016]. Additionally, "soft computing", such as artificial neural networks (ANNs) and fuzzy logic (FL), tend to be more efficient and less time consuming in the modelling of complex systems [Pahlavan *et al.* 2012; Sonmez *et al.* 2018]. Many reports discuss the efficiency of ANNs as a data driven technique to model and to forecast water consumption [Al-Zahrani, Abo-Monasar 2015; Bennett *et al.* 2012; Ghiassi *et al.* 2017; Rangel *et al.* 2017; Sakaa, Hani 2020]. A lot of work has been done in the subject. Pulido-Calvo *et al.* [2003] forecasted the total daily water demand in Fuente Palmera, Spain. Al-Zahrani and Abo-Monasar [2015] predicted the daily water demand in Al-Khobar, Saudi Arabia. Moreover, studies carried out by Adamowski *et al.* [2012] confirmed the reliability of results obtained from ANNs than other statistical methods.

Until now, little importance has been given to water consumption in Algeria, and no studies have been done on determinants of household water consumption. A better understanding of factors affecting water use and consumers behaviours can help to satisfy water demand through planning. This paper considers the assessment of domestic water use in the city of Sedrata, North-East Algeria.

MATERIAL AND METHODS

The main objectives have been to:

- determine the possible association between variables,
- assess the overall impact of different factors on household water use (PCA and ANNs),
- create and chose the best model to predict the household water consumption in the Sedrata City, and
- compare results with other works and explore implications for further research.

STUDY AREA

The present study was carried out in Sedrata, Souk Ahras Province, North-East Algeria. In 2017, the total population was 54,205 [ADE 2017]. The area has a warm climate with average temperature of 14.2°C and 523 mm precipitation (Fig. 1).

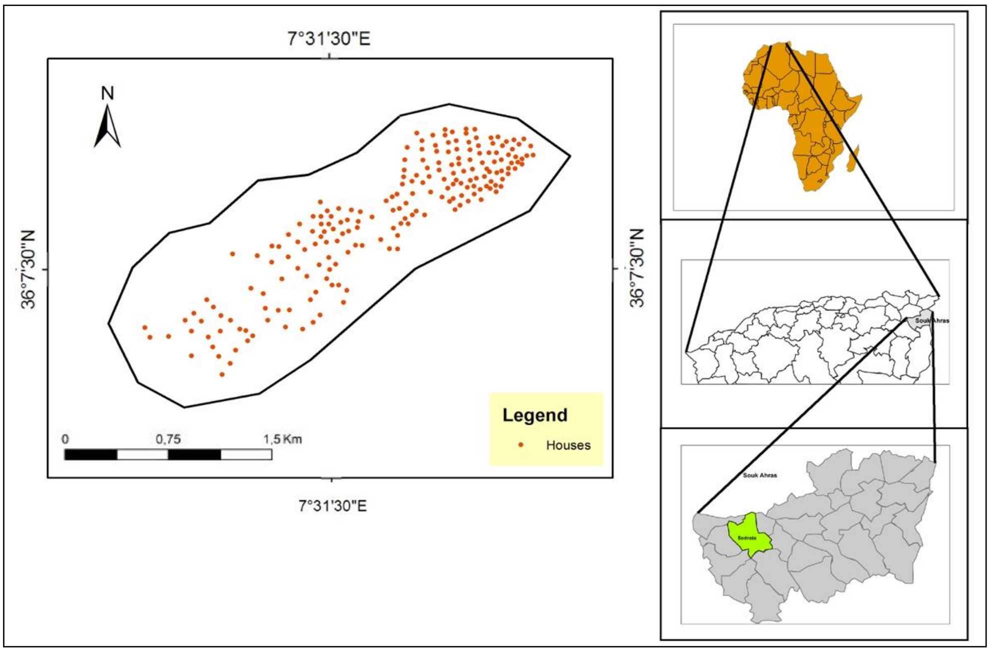


Fig. 1. Location map of study area in Sedrata City; source: own elaboration

DATA SETS

The study distinguishes three types of consumers. However, it focuses solely on the domestic use for the purpose of analysis, since it represents 89.17% of the total water consumption (ADE). Generally, the building type affects directly consumer's habits and consequently water usage. For this reason, the investigation focuses only on single houses consumers.

The identification of principal factors affecting urban water consumption is the main purpose in this paper. Based on the literature review, both quantitative and qualitative data were collected from different sources. Water consumption (WCP) is the dependent variable obtained from "Algerian water (ADE)", a water authority responsible for water distribution and management, Souk Ahras, in 2012–2017.

Other variables (dependents) were obtained by a public inquiry (Tab. 1).

Table 1. Data types and sources

		Type of data	Source
Independent variables	socio-economic parameters (SEP)	household size	public inquiry
		sex distribution	
		age categories (life cycle groupings)	
		education level of residents	
		household income	
		car possession	
	physical characteristics of building (PHC)	total area of the house	
		building area	
		number of rooms	
		garden area	
	indoor habits (INH)	clothes wash frequency	
		dishwashing frequency	
		toilets use frequency	
		shower frequency	
Dependent	Water consumption		ADE

Source: own elaboration.

The proposed methodology to assess water consumption determinants is presented in Figure 2. It attempts to better understand the relationship between domestic water use and other factors.

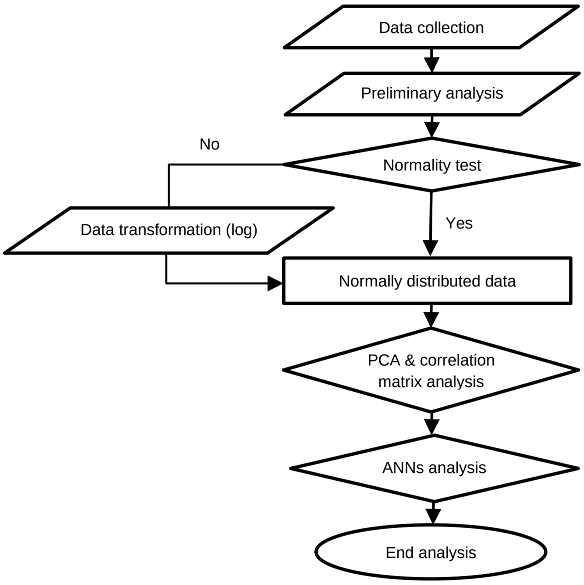


Fig. 2. The proposed methodology; source: own elaboration

DATA ANALYSIS

Preliminary data analysis. Preliminary analysis of the dataset is a critical step required before any statistical analysis. It consists of excluding outliers from rough data, unexplained noise and incomplete datasets [XUE *et al.* 2017]. The steps are shown in Figure 3.

Data preparation. Generally, the parametric analysis is preferred in every statistical analysis. However, parametric tests are mostly based on the assumption that the data distribution must be "normal". Nowadays, many non-parametric tests exist to check the normal distribution of variables, e.g. Kolmogorov-Smirnov or Shapiro–Wilk tests. The present research focuses on the parametric test, i.e. skewness and kurtosis values. A normal distribution produces a skewness and kurtosis close to "zero".

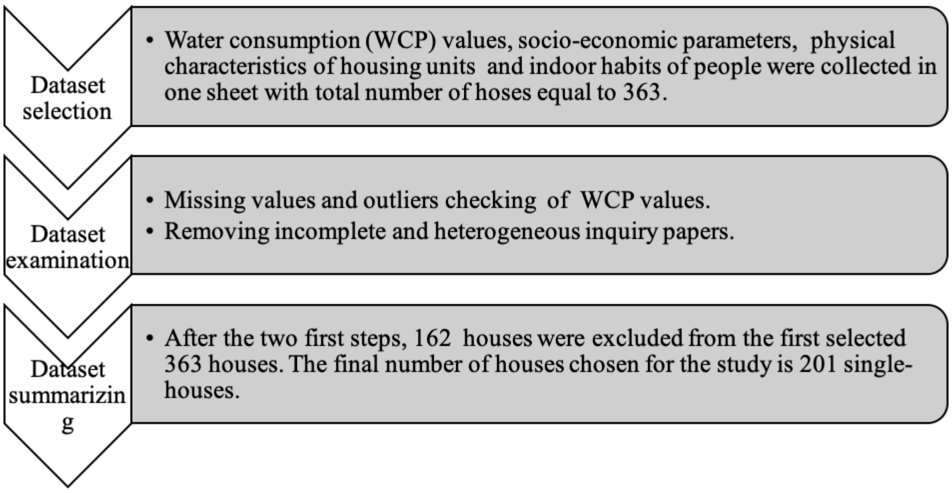


Fig. 3. Preliminary data analysis; source: own elaboration

The Principal Component Analysis (PCA). The PCA has the advantage of reducing the variables into a smaller number of factors and then sorts variables and clusters of observations with similar characteristics with respect to these factors. Additionally, it eliminates irrelevant explanatory variables. For conducting the PCA, a sample size of data must be large enough to allow correlations to converge into mutually exclusive "factors". Moreover, in this approach, normality, linearity and homogeneity of datasets are required.

Artificial Neural Networks (ANNs). Nowadays, the artificial neural networks (ANNs) are some of the most frequently used machine learning models. ANNs was first introduced in 1943 [MCCULLOCH, PITTS 1943]. The model is based on several layers containing an activation function, including the input layer (independent), hidden layers (one or more hidden layers) and the output layer (dependent). The present paper uses the multiplayer perception neural network (MLP) model.

By learning underlying relationships, ANNs can reduce development time even if the relationships are difficult to identify. Moreover, they can likewise adapt for different situations. Since generalization is needed in for the practical application of statistics, especially in the case of noisy and/or imperfect or incomplete data, ANNs can handle new data that only resemble data trained and measure the fault tolerance. ANNs can define complex links among input variables in a database. They are highly parallel, *i.e*. independent operations can be executed concomitantly [AGGRAWAL, SONG 1997].

ANNs can generalize and effectively handle information that just resembles the information the networks were initially prepared to deal with. They can also deal with defective or fragmented information due to the adaptation to non-critical failure. Speculation is especially valuable in real applications, since certifiable information is not always available. ANNs can catch complex relations between information-related factors in a network.

Although it is computing resources and time consuming, ANNs have other shortcomings, e.g. it is difficult to use and interpret findings and in some cases these are inexplicable. In addition, training methods are not yet fully understood, whereas a few methodological rules exist to establish the optimum architecture, and there is no clear way of choosing the right variant which also depends on practice and the accuracy of the training data set [AGGRAWAL, SONG 1997]. In this paper, the back-propagation feed-forward MLP with sigmoidal-type activation function (Eq. 1) has been applied as it is the most popular neural network architecture in use due to its high performance compared to other networks [LIPPMANN 1987].

$$f(z) = \frac{1}{1+\exp^{-z}} \tag{1}$$

The performance function is one of the important factors that affect the learning performance. In the feed forward network, the mean squared error (*MSE*) is commonly used as the performance function. It calculates cumulative values between target outputs and outputs created by the network.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[e(t)]^2 \tag{2}$$

where: $e(t)$ = forecast error at period t, n = number of periods [AGAMI *et al*. 2009].

RESULTS AND DISCUSSION

DESCRIPTIVE DATA

The total water consumption in Sedrata has been increasing over the last years. The evolution of water consumption (WCP) in 2012–2017 is shown in Figure 4. The increase in the WCP is due to a number of factors, for instance the population growth. The present paper focuses on 2017 water consumption and factors that influence the use of water.

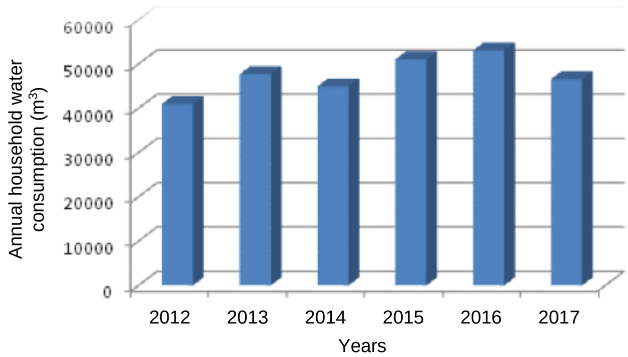


Fig. 4. Annual water consumption from 2012 to 2017; source: own study

Table 2 shows the statistic description of variables. More than 20 parameters are used to assess factors determining the domestic water use.

Essentially, four main statistical parameters are given in Table 2. These include the mean, standard deviation, skewness and the kurtosis. Results show that all the parameters had skewness and kurtosis values close to zero, which indicates a reasonably normal data distribution. In addition, only garden areas show larger positive skewness and kurtosis reflecting the non-normal distribution of data. Furthermore, results are confirmed graphically by using normal and lognormal probability plots. As a result, transformation using logarithms is only required for garden area values to produce symmetrical data.

PRINCIPAL COMPONENT ANALYSIS (PCA) RESULTS: EXPLANATORY VARIABLES

Variables are classified into three groups: the socio-economic parameters, physical characteristics of housing units, and the indoor habits. The first group (socio-economic parameters) includes 14 variables. The correlation between principal components and original variables has been shown in Figure 5. Therefore, three factors were chosen for analysis with the cumulative variance of 77.47%. Results from Figure 5 illustrate that factor F1 is determined by water consumption (WCP), number of women (FEM), household size (HOUS), two age categories: below 8 years old (AG1) and between 15 to 35 years old (AG3), and three education level of residents: primary school (PRS), high school (HGS) and university (UNIV), monthly income (INC) and number

Table 2. Statistical description of variables

Variable	Acronym	Min	Max	Mean	Standard deviation	Skewness	Kurtosis
Household water consumption (m^3)	WCP	6	75	30.09	17.19	0.62	–0.30
SOCIO-ECONOMIC INDICATOR (SEP)							
Family composition and gender							
Household size	HOUS	2	8	5	1.62	–0.37	–0.71
Number of female	FEM	1	6	3	1.26	0.24	–0.51
Number of male	MAL	0	5	2	0.89	0.38	–0.16
Age categories							
Under 8 years old	AG1	0	3	1	0.78	0.05	–0.47
Between 9 to 15 years old	AG2	0	2	1	0.67	0.51	–0.73
Between 15 to 35 years old	AG3	1	4	2	1.12	0.46	–1.21
Older than 35 years old	AG4	0	2	1	0.71	–0.32	–0.98
Education level							
Primary school	PRS	0	2	1	0.6	0.73	–0.44
Medium school	MDS	0	2	1	0.7	0.22	–0.78
High school	HGS	1	3	2	0.8	0.81	–0.86
University	UNIV	0	2	1	1.0	0.08	–0.11
Household income (DA)	INC	35 000	110 000	53 905.47	296.48	1.09	–0.13
Car possession							
Number of cars	CARN	0	3	1	0.69	–0.33	–0.72
Washing cars frequency (month)	WCAR	0	4	1	1.19	0.85	–0.05
PHYSICAL CHARACTERISTICS OF BUILDINGS (PHC)							
Total area (m^2)	TAR	80	320	187	77.29	0.15	–1.27
Building area (m^2)	BAR	40	302	165	75.81	0.19	–1.19
Number of rooms	ROMN	2	13	6	3.09	0.92	–0.20
Garden possession							
Garden area	GAR	2	80	21.7	18.14	**2.16**	**4.11**
Garden watering frequency	GWAT	1	4	2	0.79	0.56	–0.22
INDOOR HABITS (INH)							
Clothes wash frequency (week)	WCL	1	4	2	0.69	0.84	0.86
Dishwashing frequency (day)	WDISH	1	3	3	0.61	–1.13	0.24
Toilets use frequency (day)	UTLT	3	7	4	0.95	0.61	0.37
Shower frequency for female (week)	FSHW	1	7	2	1.00	0.58	1.16
Shower frequency for male (week)	MSHW	1	5	2	0.95	0.52	–0.52

Source: own study.

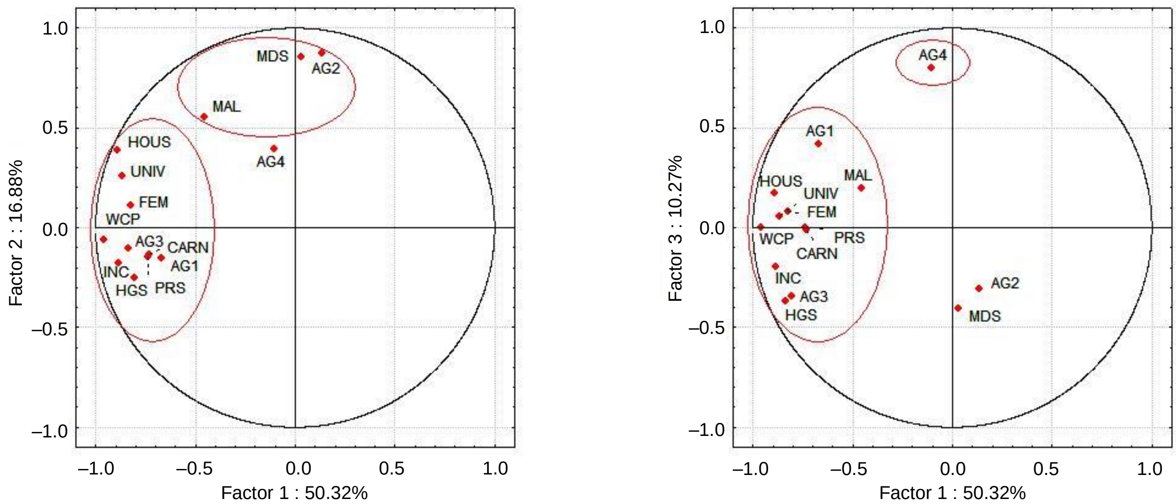


Fig. 5. Projection of variables on the factor-plane (1×2) and (1×3) for socio-economic parameters; variables acronyms as in Tab. 2; source: own study

of cars (CARN). F1 could explain factors influencing water consumption. Factor F2 is determined by MAL, AG2 and MDS. F2 represents the young male category. Factor F3 is determined by AG4 and represents the older residents category.

Physical characteristics of housing units (second group) contain 6 variables. The correlation between principal components and original variables is illustrated in Figure 6. Additionally, two factors are chosen for analysis with cumulative variance of 86.79%. This means that these two factors

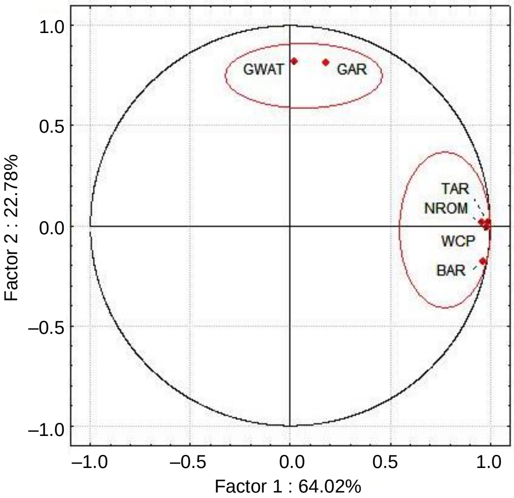


Fig. 6. Projection of variables on the factor-plane (1×2) for physical characteristics of housing units; variables acronyms as in Tab. 2; source: own study

produce sufficient information to conduct the study. Figure 6 shows that factor F1 is determined by water consumption (WCP), total area (TAR), building area (BAR) and number of rooms (NROM). F1 can be explained by household water usage. The second factor of F2 is determined by GAR and GWAT, and it represents the garden area variation.

The last group (indoor habits of residents) has 7 variables. Figure 7 shows the correlation between the principal components and the original variables. Three factors were chosen for analysis with cumulative variance of 60.90%. F1 contains washing cars (WCAR), shower for female (FSHW), and shower for male (MSHW). This factor can explain the variation of shower and car washing frequency. The second factor, F2 is determined by WCP, WCL and UTLT. It represents personal water use. The last factor, F3 is determined by WDISH. It explains dish washing frequency variation.

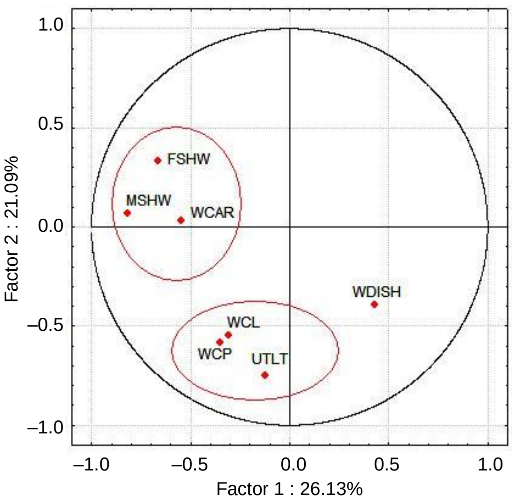
Considering that water consumption is the variable to be explained, factor F1 also controls 12 other variables, as shown in Table 3. The analysis of the correlation matrix shows that these 12 variables, TAR, BAR, NROM, FEM, HOUS, AG1, AG3, PRS, HGS, UNIV, INC, and CARN, are strongly correlated with each other.

This means that they provide the same information. Furthermore, water consumption (WCP) has different correlation coefficients with the other variables (0.96 with TAR, 0.94 with BAR, 0.93 with NROM, 0.78 with FEM, 0.81with HOUS, 0.66 with AG1, 0.76 with AG3, 0.73 with PRS, 0.78 with HGS, 0.79 with UNIV, 0.90 with INC, and 0.69 with CARN. Variables such as total area of the house, building area, number of rooms and monthly income had high correlation coefficient (>0.90). As a result, the chosen explanatory variables for this study are: water consumption, total area of the house, usable area, number of rooms, number of female, household size, two age categories of residents (AG1 and AG3), three categories of education level (primary school, high school and university), and monthly income and the number of cars.

ANNS RESULTS: MODEL ARCHITECTURES AND THEIR PERFORMANCE

Architectures of the obtained neural forecast models are summarized in Table 4. These models show the most optimal performance and they have a minimum number of hidden neurons. The table also summarizes the performance of 5 models selected for the purpose of this article. The first model (M1) has nine inputs which represent socio-economic parameters: FEM, HOUS, AG1, AG3, PRS, HGS, UNIV, INC and CARN. The next two models (M2 and M3) have three inputs: TAR, BAR and NROM, representing physical characteristics of housing units. The last two models (M4 and M5) are the combination of socio-economic parameters and physical characteristics of housing units. They contain twelve inputs (all dataset).

According to the analysis, five models attract attention since they have a correlation coefficient larger than 0.95 for all phases, i.e., they are more efficient in forecasting water consumption in Sedrata. For the following two models (12 4 1) and (12 7 1), the correlation between calculated (predicted) consumption and the observed one is fairly



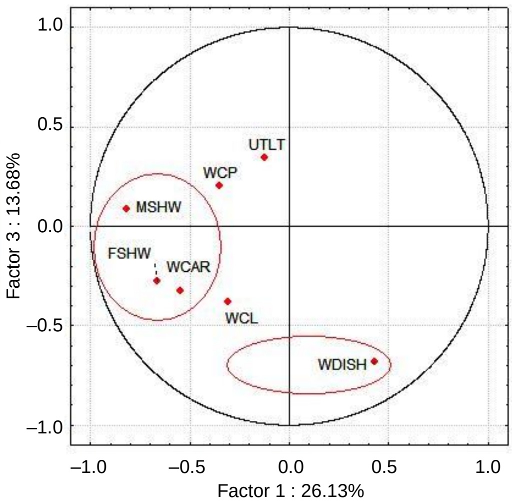


Fig. 7. Projection of variables on the factor-plane (1×2) and (1×3) for indoor habits of residents; variables acronyms as in Tab. 2; source: own study

Table 3. Variables correlation matrix

Variable	WCP	TAR	BAR	NROM	GAR	GWAT	FEM	MAL	HOUS	AG1	AG2	AG3	AG4	PRS	MDS	HGS	UNIV	INC	CARN	WCAR	WDISH	WCL	UTLT	FSHW	MSHW
WCP	1.00																								
TAR	**0.96**	1.00																							
BAR	**0.94**	**0.97**	1.00																						
NROM	**0.93**	**0.93**	**0.90**	1.00																					
GAR	0.17	0.20	–0.04	0.18	1.00																				
GWAT	0.00	0.02	–0.06	0.02	0.35	1.00																			
FEM	**0.78**	**0.73**	**0.73**	**0.71**	0.06	–0.08	1.00																		
MAL	0.37	0.43	0.41	0.31	0.09	0.11	0.11	1.00																	
HOUS	**0.81**	**0.81**	**0.80**	**0.72**	0.10	0.00	**0.84**	**0.63**	1.00																
AG1	**0.66**	**0.60**	**0.59**	**0.52**	0.10	–0.04	**0.60**	0.30	**0.63**	1.00															
AG2	–0.16	–0.14	–0.13	–0.13	–0.08	–0.01	–0.02	0.32	0.16	–0.30	1.00														
AG3	**0.76**	**0.82**	**0.79**	**0.75**	0.18	0.01	**0.64**	0.32	**0.67**	0.33	–0.20	1.00													
AG4	0.08	0.02	0.04	0.01	–0.09	0.03	0.26	0.31	0.38	0.11	0.07	–0.21	1.00												
PRS	**0.73**	**0.71**	**0.69**	**0.68**	0.15	0.02	0.48	0.35	**0.57**	**0.60**	–0.15	**0.54**	–0.07	1.00											
MDS	–0.06	0.00	–0.01	–0,05	0.02	0.05	0.12	0.27	0.25	–0.22	**0.84**	0.03	–0.03	–0.07	1.00										
HGS	**0.78**	**0.75**	**0.75**	**0.81**	0.09	–0.08	**0.61**	0.16	**0.56**	0.31	–0.20	**0.83**	–0.17	**0.58**	–0.18	1.00									
UNIV	**0.79**	**0.82**	**0.80**	**0.73**	0.13	–0.02	**0.76**	**0.52**	**0.87**	0.44	0.02	**0.75**	0.30	**0.50**	0.15	**0.63**	1.00								
INC	**0.90**	**0.86**	**0.85**	**0.87**	0.12	–0.06	**0.68**	0.24	**0.66**	0.46	–0.18	**0.77**	–0.05	**0.61**	–0.14	**0.88**	**0.72**	1.00							
CARN	**0.69**	**0.70**	**0.69**	**0.62**	0.13	–0.08	**0.50**	0.32	**0.57**	**0.55**	–0.16	**0.57**	–0.05	**0.53**	–0.12	**0.53**	**0.53**	**0.64**	1.00						
WCAR	0.10	0.12	0.08	0.12	0.14	0.07	0.07	0.08	0.10	0.02	–0.06	0.12	0.07	–0.02	–0.07	0.10	0.13	0.12	0.12	1.00					
WDISH	0.02	–0.02	0.01	0.01	–0.01	0.15	0.04	–0.06	0.00	–0.02	–0.07	0.03	0.04	0.03	–0.07	0.07	0.04	0.03	–0.07	–0.06	1.00				
WCL	0.16	0.17	0.15	0.11	0.13	0.21	0.03	0.19	0.13	0.15	–0.07	0.08	0.08	0.13	–0.04	0.04	0.18	0.12	0.12	0.06	0.10	1.00			
UTLT	0.28	0.28	0.29	0.23	–0.02	0.03	0.20	0.14	0.23	0.08	0.01	0.24	0.05	0.17	0.05	0.24	0.29	0.28	0.20	0.00	0.06	0.22	1.00		
FSHW	0.04	0.06	0.03	0.08	0.10	0.01	–0.03	0.02	–0.01	–0.02	–0.01	0.05	0.01	0.03	–0.13	0.05	0.02	0.06	0.13	0.20	–0.18	0.08	–0.16	1.00	
MSHW	0.16	0.19	0.14	0.17	0.26	0.15	0.08	0.08	0.10	0.02	–0.13	0.19	0.03	0.06	–0.11	0.13	0.16	0.18	0.21	0.29	–0.30	0.15	0.08	0.42	1.00

Explanations: variables acronyms as in Tab. 2.
Source: own study.

Table 4. Architectures and the performance of neural models

Model	Input	Structure	Hidden layer	Training MSE	Validation MSE	Testing MSE	All MSE	Training R	Validation R	Testing R	All R
M1	S1	(9 7 1)	7	5.76	11.86	13.84	**9.28**	0.99	0.98	0.98	**0.98**
M2	S2	(3 4 1)	4	15.26	7.41	5.62	**8.71**	0.97	0.98	0.98	**0.97**
M3	S2	(3 4 1)	4	11.55	5.72	1.64	**8.82**	0.98	0.99	0.99	**0.98**
M4	S1+S2	(12 4 1)	4	1.37	1.84	1.34	**1.41**	0.99	0.99	0.99	**0.99**
M5	S1+S2	(12 7 1)	7	0.27	3.01	1.89	**1.19**	0.99	0.99	0.99	**0.99**

Explanations: M1–M5 = model from 1 to 5; S1 = socio-economic parameters (S1 = FEM + HOUS + AG1 + AG3 + PRS + HGS + UNIV + INC + CARN), S2 = physical characteristics of housing units (S2 = TAR + BAR + NROM); variables acronyms as in Tab. 2.
Source: own study.

strong (0.99): these models are therefore very efficient. The mean square error (*MSE*) of the consumption was also analysed, especially during the training phase. It should be remembered that lower values of the *MSE* are better and the model is well optimized if the values are close to zero. During the training phase, the *MSE* coefficient for various models (12 4 1) and (12 7 1) is the smallest, respectively 1.37 and 0.27, while *MSE* values are 5.76, 15.26 and 11.55 respectively for (9 7 1), (3 4 1) and (3 4 1) models. Taking these results into account, the rest of the study focuses on models (12 4 1) and (12 7 1), which appear to be the most efficient for forecasting domestic water consumption in Sedrata. During the testing phase, correlation coefficient values in the selected models are similar and equal to 0.99. *MSE* values obtained during the test phase also allow the two models to be compared. The *MSE* of the (12 4 1) model is smaller than the error of the (12 7 1) model. We could observe that the correlation coefficients (*R*) of the first three models (9 7 1), (3 4 1) and (3 4 1) are higher but the *MSE* values (Fig. 8) are larger compared with (12 4 1) and (12 7) models. The analysis of results indicates that (12 4 1) and (12 7 1) models better predict water consumption of Sedrata.

The architecture of (12 4 1) and (12 7 1) models is presented in Figure 9, including 1 hidden layer with 4 and 7 neurons, respectively.

The comparison (Fig. 10) between the calculated or predicted consumption and the observed one reveals that models (12 4 1) and (12 7 1) could predict domestic water consumption very well. Moreover, the two groups of variables (socio-economic parameters and physical characteristics of building units) combined a lot more information to neural models developed in this study than separately (models (9 7 1), (3 4 1) and (3 4 1)) to predict household water consumption in Sedrata. We can deduce that these variables induce a positive effect on water consumption in the region. In addition, correlation coefficients are equal to 0.99 in the training, testing and validation phases for the two models. In general, the combination of variables significantly

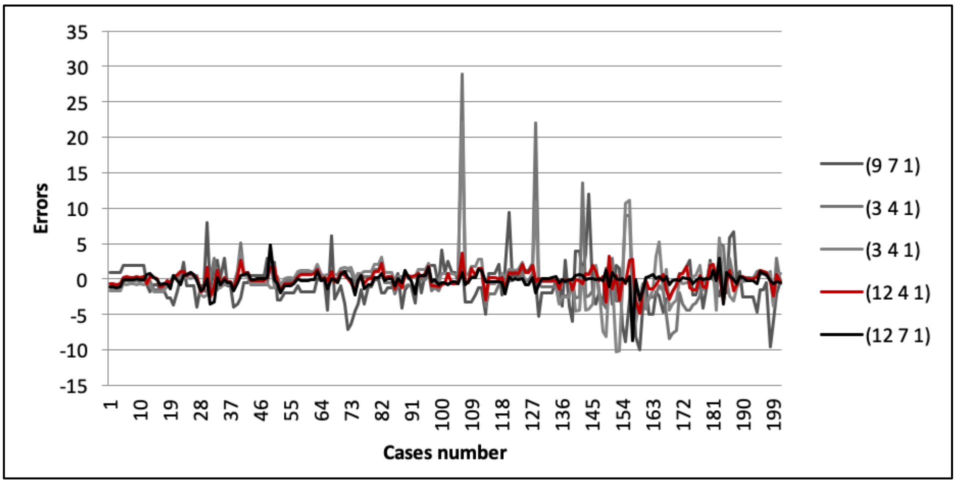


Fig. 8. Comparison of errors in the five selected models; source: own study

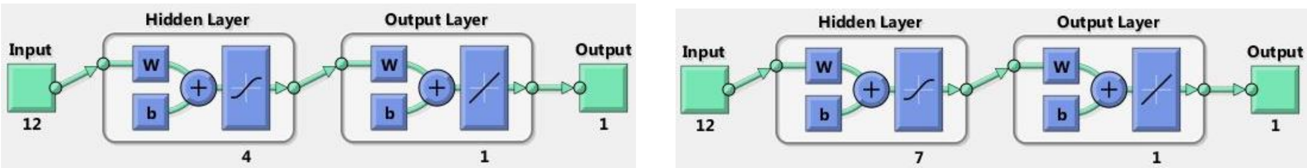


Fig. 9. Neural network structures of the two selected models; source: own study

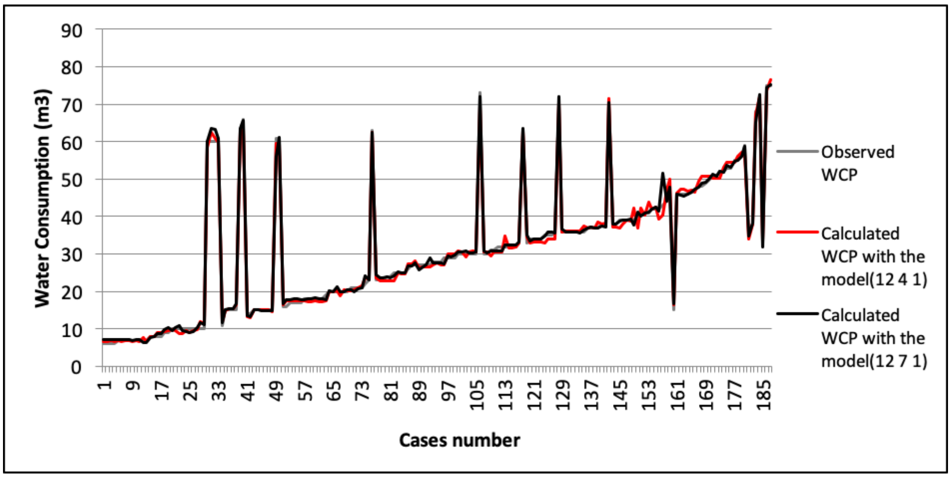


Fig. 10. Comparison between observed and calculated (predicted) water consumption (WCP) in (12 4 1) and (12 7 1) models; source: own study

improves model performance, i.e. the choice of input variables greatly influences the performance of neural networks. Other studies which included the same principal components but different explanatory variables produced better results like AL-ZAHRANI and ABO-MONASAR [2015].

CONCLUSIONS AND RECOMMENDATIONS

The study employed the principal component analysis to evaluate the relative influence of several parameters (i.e. socio-economic parameters and physical characteristic of buildings) on domestic water consumption. This has been based on the case study of Sedrata, Algeria. Results show that some of the two last categories of variables are strongly correlated with water consumption. Additionally, the artificial neural networks technique was applied to assess determinants of household water use. It is found that twelve (12) variables are the dominant drivers in household water use, e.g. household size, monthly income of residents, and area of the house.

As for the impact of socio-economic parameters, the results show that the number of women, household size, two age categories of residents (AG1 and AG3), three categories of education level (primary school, high school and university), monthly income and the number of cars in a household have a significant impact on water consumption, while the number of men, two age categories of residents (AG2 and AG4) and education level (medium school) have no influence on the use of water. As regards physical characteristics of housing units, the total area of the house, building area and the number of rooms have a significant influence on water consumption, whereas the garden area had no effect on water use.

Results from the study could help any water authority to develop effective strategies designed to satisfy water needs in urban areas. The future water demand is expected to rise especially with the global warming and the increase in population. The findings of the present study highlight the significance of physical and socio-economic factors over climatic factors in the forecasting of water consumption. However, climatic factors must be also considered while assessing water usage to link the probable impact of climate change on water demand.

REFERENCES

ADAMOWSKI J., CHAN H.F., PRASHER S.O., OZGA-ZIELINSKI B., SLIUSARIEVA A. 2012. Comparison of multiple linear and nonlinear regression, Autoregressive Integrated Moving Average, Artificial Neural Network, and Wavelet Artificial Neural Network Methods for Urban Water Demand Forecasting in Montreal, Canada. Water Resources Research. Vol. 48. Iss. 1, W01528 p. 1–14. DOI 10.1029/ 2010WR009945.

ADE 2017. Water authority of Souk-Ahras Province. Algérienne des eaux. [Access 15.04.2020]. Available at: https://www.ade.dz/index.php/espace-client/services

AGAMI N., ATIYA A., SALEH M., EL-SHISHINY H. 2009. A neural network based dynamic forecasting model for trend impact analysis. Technological Forecasting and Social Change. Vol. 76. Iss. 7 p. 952–962. DOI 10.1016/j.techfore.2008.12.004.

AGGRAWAL R., SONG Y. 1997. Artificial neural networks in power systems. Part 1. General introduction to neural computing. Power Engineering Journal. Vol. 11. Iss. 3 p. 129–134. DOI 10.1049/pe:19970306.

AL-ZAHRANI M.A., ABO-MONASAR A. 2015. Urban residential water demand prediction based on artificial neural networks and time series models. Water Resources Management. Vol. 29 p. 3651–3662. DOI 10.1007/s11269-015-1021-z.

ARBUÉS F., VILLANUA I., BARBERAN R. 2010. Household size and residential water demand: An empirical approach. The Australian Journal of Agricultural and Resource Economics. Vol. 54. Iss. 1 p. 61–80. DOI 10.1111/j.1467-8489.2009. 00479.x.

Aquacraft Inc. 2015. Application of end use study data for development of residential demand models [online] pp. 3. [Access 15.04.2020]. Available at: http://www.aquacraft.com/wp-content/uploads/2015/09/Residential-Models.pdf

BALLING Jr. R.C., GOBER P., JONES N. 2008. Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. Water Resources Research. Vol. 44. Iss. 10, W10401 p. 1–11 DOI 10.1029/ 2007WR006722.

BEAL C., STEWART R.A., HUANG T.T., REY E. 2011. SEQ residential end use study. Water (Australia). Vol. 38. Iss. 1 p. 80–84.

BENNETT C., STEWART R.A., BEAL C.D. 2012. ANN-Based residential water end-use demand forecasting model. Expert Systems with Application. Vol. 40. Iss. 4 p. 1014–1023. DOI 10.1016/j.eswa.2012.08.012.

FAN L., GAI L., TONG Y., LI R. 2017. Urban water consumption and its influencing factors in China: Evidence from 286 cities. Journal of Cleaner Production. Vol. 166 p. 124–133. DOI 10.1016/j.jclepro.2017.08.044.

FIELDING K.S., RUSSELL S., SPINKS A., MANKAD A. 2012. Determinants of household water conservation: The role of demographic, infrastructure, behavior and psychosocial variables. Water Resources Research. Vol. 48. DOI 10.1029/2012 WR012398.

GATO-TRINIDAD S., JAYASURIYA N., ROBERTS P. 2011. Understanding urban residential end uses of water. Water Science and Technology. Vol. 64. Iss. 1 p. 36–43. DOI 10.2166/ wst.2011.436.

GHIASSI M., FA'AL F., ABRISHAMCHI A. 2017. large metropolitan water demand forecasting using DAN2, FTDNN, and KNN Models: A case study of the city of Tehran, Iran. Urban Water Journal. Vol. 14. Iss. 6 p 655–659. DOI 10.1080/ 1573062X.2016.1223858.

GRAFTON R.Q., WARD M.B., TO H., KOMPAS T. 2011. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. Water Resources Research. Vol. 47. Iss. 8. DOI 10.1029/ 2010WR009685.

HAQUE MD M., EGODAWATTA P., RAHMAN A., GOONETILLEKE A. 2015. Assessing the significance of climate and community factors on urban water demand. International Journal of Sustainable Built Environment. Vol. 4. Iss. 2 p. 222–230. DOI 10.1016/j.ijsbe.2015.11.001.

HOUSE-PETERS L.A., CHANG H. 2011. Urban water demand modeling: Review of concepts, methods, and organizing principles. Water Resources Research. Vol. 47. Iss. 5, W05401 p. 1–15. DOI 10.1029/2010WR009624.

HUSSIEN W.A., MEMON F.A., SAVIC D.A. 2016. Assessing and modelling the influence of household characteristics on per capita water consumption. Water Resources Management. Vol. 30 p. 2931–2955. DOI 10.1007/s11269-016-1314-x.

KENNEY D.S., GOEMANS C., KLEIN R., LOWREY J., REIDY K. 2008. Residential water demand management: Lessons from Aurora, Colorado. Journal of the American Water Resources Association. Vol. 44. Iss. 1 p. 192–207. DOI 10.1111/j.1752-1688.2007.00147.x.

LIPPMANN R.P. 1987. An introduction to computing with neural nets. IEEE ASSP Magazine. Vol. 4. Iss. 2 p. 4–22. DOI 10.1109/MASSP.1987.1165576.

MAKKI A.A., STEWART R.A., PANUWATWANICH K., BEAL C. 2011. Revealing the determinants of shower water end use consumption: Enabling better targeted urban water conservation strategies. Journal of Cleaner Production. Vol. 60 p. 129–146. DOI 10.1016/j.jclepro.2011.08.007.

MAKKI A.A., STEWART R.A., BEAL C.D., PANUWATWANICH K. 2015. Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. Resources, Conservation and Recycling. Vol. 95 p. 15–37. DOI 10.1016/j.resconrec.2014.11.009.

MATOS C., TEIXEIRA C.A., BENTO R., VARAJÃO J., BENTES I. 2014. An exploratory study on the influence of socio-demographic characteristics on water end uses inside buildings. Science of The Total Environment. Vol. 466–467 p. 467–474. DOI 10.1016/j.scitotenv.2013.07.036.

MCCULLOCH W.S., PITTS W.H. 1943. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics. No. 5 p. 115–133. DOI 10.1007/BF02459570.

MU X., WHITTINGTON D., BRISCOE J. 1990. Modeling village water demand behavior: A discrete choice approach. Water Resources Research. Vol. 26. Iss. 4 p. 521–529. DOI 10.1029/ WR026i004p00521.

NAUGES C., WHITTINGTON D. 2009. estimation of water demand in developing countries: An overview. The World Bank Research Observer. Vol. 25. Iss. 2 p. 263–294. DOI 10.1093/wbro/ lkp016.

PAHLAVAN R., OMID M., AKRAM A. 2012. Application of data envelopment analysis for performance assessment and energy efficiency improvement opportunities in greenhouses cucumber production. Journal of Agricultural Science and Technology. Vol. 14 p. 1465–1475.

PULIDO-CALVO, J.R., LOPEZ-LUQUE R., GUTIERREZ-ESTRADA J.C. 2003. Demand forecasting for irrigation water distribution systems. Journal of Irrigation and Drainage Engineering. Vol. 129. Iss. 6 p. 422–431. DOI 10.1061/(ASCE)0733-9437(2003)129.

RANGEL H.R., PUIG V., FARIAS R.L., FLORES J.J. 2017. Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks. Journal of Hydroinformatics. Vol. 19. Iss. 1 p. 1–16. DOI 10.2166/hydro.2016.199.

SAKAA B., CHAFFAI H., HANI A. 2020. ANNs approach to identify water demand drivers for Saf-Saf river basin. Journal of Applied Water Engineering and Research. Vol. 8. Iss. 1 p. 44–54. DOI 10.1080/23249676.2020.1719220.

SONMEZ A.Y., KALE S., OZDEMIR R.C., KADAK A.E. 2018. An adaptive neuro-fuzzy inference system (ANFIS) to predict of cadmium (Cd) concentrations in the Filyos River, Turkey. Turkish Journal of Fisheries and Aquatic Sciences. Vol. 18. Iss. 12 p. 1333–1343. DOI 10.4194/1303-2712-v18_12_01.

TIWARI M.K., ADAMOWSKI J. 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resources Research. Vol. 49. Iss. 10 p. 6486–6507. DOI 10.1002/wrcr.20517.

TIWARI M., ADAMOWSKI J., ADAMOWSKI K. 2016. Water demand forecasting using extreme learning machines. Journal of Water and Land Development. No. 28 p. 37–52. DOI 10.1515/jwld-2016-0004.

UNESCO 2012. Global water resources under increasing pressure from rapidly growing demands and climate change, according to new UN World Water Development Report [online]. [Access 15.04.2020]. Available at: http://www.unesco.org/new/en/media-services/single-view/news/global_water_resources_under_increasing_pressure_from_rapidl/

WENTZ E.A., WILLS A.J., KIM W.K., MYINT S.W., GOBER P., BALLING JR R.C. 2014. Factors influencing water consumption in multifamily housing in Tempe, Arizona. The Professional Geographer. Vol. 66. Iss. p. 501–510. DOI 10.1080/00330124.2013.805627.

WILLIS R., STEWART R.A., PANUWATWANICH K., CAPATI B., GIURCO D. 2009. Gold coast domestic water end use study. Water. Vol. 36. Iss. 6 p. 79–85.

XUE P., HONG T., DONG B., MAK C. 2017. A preliminary investigation of water usage behavior in single-family homes. Building Simulation. Vol. 10 p. 949–962. DOI 10.1007/ s12273-017-0387-7.