Urban domestic water consumption behavior: a case study in Cuiabá city, Mato Grosso state, Brazil

W. T. P. Silva[1,2]; V. S. M. Silva[3]; L. A. Gomes[1] and M. M. Espinosa[3]
[1] Department of Sanitary and Environmental Engineering, Faculty of Architecture, Engineering and Technology, Federal University of Mato Grosso, Cuiabá, Brazil. E-mail: wttatom@terra.com.br (W. T. P. Silva), la.gomes@terra.com.br (L. A. Gomes)
[2] Post-graduate Program on Environmental Technology and Water Resources, Department of Civil and Environmental Engineering, Faculty of Technology, University of Brasilia, Brasilia, Brazil.
[3] Post-graduate Program on Forestry and Environmental Sciences, Department of Forestry Engineering, Faculty of Forestry Engineering, Federal University of Mato Grosso, Cuiabá, Brazil. E-mail: versides@uol.com.br (V. S. M. Silva), marianom@cpd.ufmt.br (M. M. Espinosa)

Abstract
The main aim of this study was to determine a statistical model capable of projecting a per capita water quota for the city of Cuiabá, Mato Grosso State, Brazil. The sample was composed of the 115 recognized neighborhoods of Cuiabá, separated by month and referring to the period between 2003 and 2006. The methodology used was based on the statistical model (analysis of regression and verification of the models). As a result we obtained a model stratified based on socioeconomic status with reasonable projection capability of per capita share to the Cuiabá city. A combination of grouping techniques using the models of Burtless and Hausman (1978) or artificial intelligence techniques could be an alternative for obtaining better models.

Keywords: water management, statistical modeling, water resources

Introduction
The urban sprawl is presented as an intrinsic process of urban development in Brazil. As a consequence there are impacts on the stability of the environment, producing unskilled and unhealthy places, providing exclusion, marginalization and imposing a rhythm of life incompatible with the established concepts such as quality of life (Rossetto *et al*., 2004). Regarding the impact on water resources, often there is no compromising between quality and quantity due to the discharge of sewage and high levels of losses in the water supply system, respectively. Such impacts project several scenarios in cities, among them, the poor sanitation conditions, the high cost of treating water, the water scarcity and the search for alternative sources of supply. Not far from these is the city of Cuiabá, that due to rapid urbanization and economic growth came from the 70's, it came to reach and directly affect the Cuiabá River. According to the municipal government of this city (Cuiabá, 2006), sanitation has not kept pace of city growth, which affects water quality, bank ciliary vegetation and integrity of the river bed. Still is observed that there is water loss rate of around 65% (ISA 2004) and the existence of an intermittent water supply. The Master Plan and Management System for Water Resources come as an alternative to promote the orderly growth of cities and minimize environmental impacts.

According to Silva and Porto (2003), a broader view is a vital element of the concept of integrated urban planning programs and actions for the conservation and rational use of water. Still, according the same author, this integrated approach is grounded in actions by the demand of water. In agreement, Silva and Rocha (1999) report that the forecast demand for water is a basic tool for planning, necessary for the correct sizing of supply and to direct measures of water management. Also, the Mato Grosso State Environmental Agency (SEMA) reported that the development of specific research on demands for water is an essential action to the management of groundwater and surface water in urban area of Cuiabá (MT-FEMA, 2004).

The need for actions to demand management of water, presented for the case of Cuiabá, suggests the development of instruments aimed to forecast the demand for water. Thus, this paper aims to propose a statistical model able to project the per capita share of water contributing to: (i) formulation of management tools for water resources, (ii) preparation of master plans, (iii) investment forecast in sanitation, and (iv) actions for the conservation and rational use of water.

Review of related research
As the state of the art of forecasting studies of water consumption can cite some previous studies. These studies include various forecasting methods including: multiple regression, time series analysis, artificial neural networks (ANN) and fuzzy logic (FL).

In the case of the linear regression method one can mention some works, among them: Martinez-Espiñeira (2002) who used a monthly database of northeastern Spain to estimate the demand function of domestic water tariff from linear and nonlinear. Tariff, bill, climate and the socio-demographic variables were used as explanatory variables. The results indicated total marginal price elasticity of between -0.12 and -0.17, and climatic variables significantly affect the monthly data. In parallel, a study entitled "Estimating urban residential water demand in the State of Ceará", directed by IPECE (2003) presented as explanatory variables: income, price, number of rooms and the number of residents per household. It was also noted that there are differences of demand between: households connected to the mains sewerage and water connected only. The study presented as steps of the methodology to define the sample size, the regionalization of the State of Ceará, the selection of observational units, in addition to questionnaires. Fernandes-Neto *et al*. (2004) studied several variables involved in water per capita for a sample universe consisting of 96 municipalities in Minas Gerais State. As steps towards achieving the study note: (1) the sample universe definition, (2) definition of variables of interest and (3) statistical analysis. The study enabled observations infer about the influence of factors considered, according to each age population assessed and outline a mathematical model for municipalities with a population of 50 to 100 thousand inhabitants. Karlis *et al*. (2007) inserted the heteroscedastic and multiplicity to statistical models for effecting adjustment to the data set of water consumption in the city of Athens. The results indicated that the profile of lower consumption refers to single residents (male or female), also noted problems with the residual analysis of the model developed. As to the technique of time series it highlights some studies, among them, Zhou *et al*. (2000) who conducted research to forecast the daily water consumption for the city of Melbourne with methods using time series models with different equations representing four influential factors, including the trend, seasonality, climatic correlation and autocorrelation. The model was tested using the Cross-Validation procedures and independence of the data collected between December 1, 1996 to January 31, 1997. Also along these lines, Zhou *et al*. (2002) proposed a time series model to predict the hourly consumption of water for 24 hours in advance a system of water supply in urban area of Melbourne, Australia. The model was composed of two modules: the daily and hourly. As result, it was obtained a model with 66% of variance and a standard deviation of 162L.(inhab.day)$^{-1}$.

Finally, one can cite the methods of RNA and LF work of Oshima and Kosuda (1998), who developed a prediction model for water demand time from the Chaos Theory, rebuilt by Local Fuzzy Method. Results were correlated to the measured data and came to an R^2 = 0.985. Mukhopadhyay et al. (2001) used data from the weekly consumption of 48 households in Kuwait, collected over a period of one year. Based on these data the per capita residential water was estimated in the range of 182-2018L.(inhab.day)$^{-1}$, with an average of 814L.(inhab.day)$^{-1}$. The linear regression methods and RNA were used in the study in order to fit the observed values. The results suggested a dependence of water per capita consumption, bathroom and room number in the residence, garden size, income level of the residence, atmospheric temperature, relative humidity and number of people in the residence. Results were satisfactory for the proposed objectives. Falkenberg (2005) held "Forecasting urban water consumption in the short term," which aimed to present different models for forecasting consumption in the short term by using techniques such as RNA, multiple linear regression, and Box-type models and Jenkins. The results were satisfactory.

Material and methods

For this study was considered as population (universe) the city of Cuiabá and all districts recognized by the Institute of Research and Urban Development of Cuiabá (IPDU) in 2006, stratified by month and year for the period 2003 to 2006. Therefore, 115 neighborhoods were then investigated in the city for a period of 48 months. The selection of variables to the model development was based on the suggested by Fernandes-Neto *et al*. (2004) and insertion of non-conventional economic indicators for the prediction of water consumption (per capita quota of energy, energy tariff, social energy tariff, minimum wage, inflation, dollar exchange rate and interest rate). This insertion was motivated by the recommendation of Silva *et al*. (2008), when conducting studies of the per capita share of water.

The selected variables, coding, unit of measure and data source are presented in Table 1. Regarding socio-economic class, they were subdivided into five categories: low, medium low, medium, medium high and high (Cuiabá, 2006), and because it is a qualitative variable can only be inserted under the modeling the form of Dummy Variables. In this case the socio-economic class was entered as four Dummy Variables (D2, D3, D4 and D5) binary (0 or 1). Table 2 presents socio-economic classes, households income and their Dummy variables.

As for seasonality, we observed two well-defined climatic moments, the rainy season and dry season also included qualitative variables as Dummy (D6), and the value for a rainy period and the value to 0 otherwise. TA practiced is defined in terms of consumption of home range, these being: (i) range 1: 0-30m^3.month^{-1} | rate: 1.12R$.m^{-3}, (ii) range 2: 11 - 20m^3.month^{-1} | rate: 1.68R$.m^{-3}, (iii) range 3: 21-30m^3.month^{-1} | rate:

2.80R$.$m^{-3}$, (iv) range 4: 31 - 40 m^3.$month^{-1}$ | rate: 3.70R$.$m^{-3}$, (v) track 5: above 40 m^3.$month^{-1}$ | rate: 5.56 R$.$m^{-3}$ (Sanecap, 2007).

Regarding TS, this is a benefit granted to low-income population, as ANEEL (2002), and therefore inserted into the job as an instrument of socio-economic status. The range of TS, that were inserted in the form of Dummy Variables. Data concerning the tracks were obtained from the local power utility, the Rede-Cemat. In Table 3 are found information relating to tariff and its transformation to Dummy variables.

Table 1 - Selected variables for modeling

Variable	Coding	Unit	Data Source
Water consumption per neighborhoods	CB	m^3.(neighborhood.month)$^{-1}$	Sanecap
Population by district	HB	inhabitant	IPDU
Per capita share of water	PA	L.(inhab.day) $^{-1}$	*
Average monthly temperature of air	TP	ºC	INMET
Monthly average humidity of the air	UR	%	INMET
Monthly rainfall	IP	mm.month^{-1}	INMET
Neighborhood socioeconomic class	CS	[a]	IPDU
Month in study	ME	[a]	[b]
Seasonality	SZ	[a]	IPDU
Energy consumption per quarter	CE	kWh.month^{-1}	Rede-Cemat
Per capita quota of energy	PE	kWh.(inhab.day)$^{-1}$	**
Water tariff	TA	R$.$m^{-3}$	Sanecap
Social energy tariff	TS	[a]	Rede-Cemat
Real minimum wage	SR	R$	IPEA
Index of market prices (IGP-M)	IF	%	FGV
Dollar exchange rate	DO	R$	BCB
Interest rate	TJ	%	BCB

[a]: dimensionless [b]: Time series; * PA = (CB. (HB.DM) -1) .1000, ** PE = (CE. (HB.DM) $^{-1}$); DM: number of days in month (DM; Sanitation of Capital Agency (SANECAP); Institute for Research and Urban Development of Cuiabá (IPDU); National Institute of Meteorology (INMET); Central Electricity Company of Mato Grosso (REDE-CEMAT), IPEA: Institute of Applied Economic Research (IPEA), Getúlio Vargas Fundation (FGV); Central Bank of Brazil (BCB).

Table 2 - Socio-economic classes, income and Dummy variables.

Socio-economic classes	Nominal average monthly charge per household	Dummy variable for CS			
		D2	D3	D4	D5
Low Income	under of 2,91 SM	0	0	0	0
Lower Middle Income	of 2,91 S.M. to 5,65 SM	1	0	0	0
Average Income	of 5,66 S.M. to 11,65 SM	0	1	0	0
Middle-High Income	of 11,66 S.M. to 21,94 SM	0	0	1	0
High Income	above to 21,94 SM	0	0	0	1

Source: Cuiabá (2006); Minimum Wage (SM).

Table 3 - Social tariff in the form Dummy variable.

Range of intake (kWh.month^{-1})	Rate power supply (R$.(MWh)$^{-1}$	Dummy variables for TS			
		D7	D8	D9	D10
0 – 30	107,54	0	0	0	0
30 – 80	185,09	1	0	0	0
80–100	186,25	0	1	0	0
100 – 140	279,32	0	0	1	0
140 to Up	210,39	0	0	0	1

Source: Rede-Cemat (2007)

The tools used for the development of the models were: Microsoft Excel, and SPSS software TableCurve. The development of the statistical model to predict the per capita water was structured in stages, as shown in Figure 1.

In step I all the data were divided into two sets, those for modeling (75%) and data for verification (25%), the first set covering years 2003 to 2005 and the second set year 2006. Step II refers to the descriptive statistical analysis and selection of influential variables performed using correlation analysis and preparation of scatter diagrams (PA versus explanatory variables possibly).

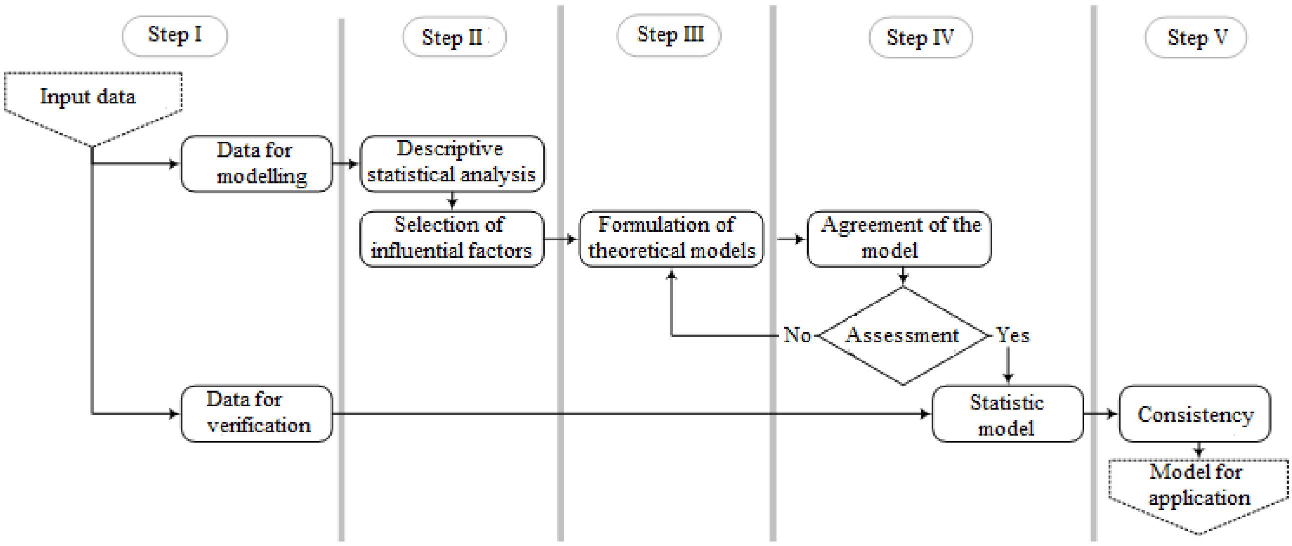


Figure 1 - Sequence of steps toward statistical modeling

From the selection of influential variables and scatter plots was held to propose a theoretical model to predict and search for non-linear models and polynomial that fit the data well (step III). An example of a linear additive model considering all the variables selected for modeling is shown in Equation 1.

$$PA = \beta_0 + \beta_1.TP + \beta_2.UR + \beta_3.IP + \beta_4.D2 + \beta_5.D3 + \beta_6.D4 + \beta_7.D5 + \beta_8.D6 + \beta_9.D7 + \beta_{10}.D8 + \beta_{11}.D9 + \beta_{12}.D10 + \beta_{13}.ME + \beta_{14}.PE + \beta_{15}.TA + \beta_{16}.SR + \beta_{17}.IF + \beta_{18}.DO + \beta_{19}.TJ + \varepsilon \quad (1)$$

Where: $\beta_0, \ldots, \beta_{19}$ = coefficients; AP = per capita share of water (L.(inhab.day)$^{-1}$); TP = temperature (°C); UR = relative humidity (%); IP = rainfall (mm.h^{-1}); D2 = Dummy medium-low class; D3 = Dummy middle class; D4 = Dummy class medium-high; Dummy D5 = upper class; D6 = Dummy seasonality; D7 = Dummy consumption range 30-80; D8 = Dummy consumption range 80-100; D9 = Dummy consumption range 100-140; D10 = Dummy range of 140-higher than consumption; ME = month; PE = energy consumption per capita (kWh.(inhab.day)$^{-1}$); TA = water tariff (R$.m^{-3}), SR = minimum wage (R$) IF = inflation (%); DO = the dollar exchange rate (R$) TJ = rate interest (%); ε = error.

In step IV was held: agreement of the model by multiple regression analysis, considering significant $p \leq 0.05$ for the remaining variable in the model and theoretical model of diagnosis through the determination coefficient (R^2), statistical F tests of homogeneity of variance (standardized residuals *versus* estimated PA), a test of normal distribution (Q-Q plots) and percentage of error (in percentage residue *versus* estimated PA). The diagnosis of non-suitability of the theoretical model (lower value of R^2, the presence of heteroscedasticity, a high percentage of waste) report the need to adapt/modify the theoretical model, for otherwise the model was said appropriate. In step V was held and the consistency test using aggregate data for verification. To this end, we are: chart observed *versus* estimated value; heteroscedastic T-test (paired samples with unequal variances), the level of significance $p \leq 0.05$, and correlation analysis between observed and predicted values. Adopted as a satisfactory observation of linearity between the observed and estimated significance level for the t-test greater than 0.05 ($p \leq 0.05$) and correlation coefficients exceeding 0.70. The procedures followed those recommended Aguirre (2007), Gujarat (2006), Souza (1998), Sokal and Rohlf (1995), Draper and Smith (1980), Neter and Wasserman (1974).

Results and discussion

Results of step II, conducted with the purpose of viewing the format and behavior of the variables are presented in Tables 4 and 5, Figures 2a, 2b and 2c, regarding statistical analysis, correlation and scatter diagram, respectively.

The descriptive statistics, Table 4, showed a mean value of PA 140L.(inhab.day)$^{-1}$, a value close to that presented by the NHIS (2004) and ISA (2007) for the region. Table 5 shows the correlation coefficients for all variables studied. We observed statistically significant associations ($p < 0.01$) between BP and socio-economic class (D2, D3, D4 and D5), social energy tariff (D7, D8, D9 D10), per capita energy use (PE) and water tariff (TA), which are the possible influential factors selected. Associations between climatic variables can be assigned to local seasonal, with two climactic moments (the rainy season and dry season) are well defined in region (Piaia, 1997). Another point refers to the associations found between climatic variables (TP, RH, and IP) and economic indicators (SR, OF, IF and TJ), which may refer to the spurious correlations, since both variables are influenced in time. The significant correlations between these variables and month (ME), seasonality (D6) and behavior of these variables over time (time series) are signs the acceptance of spurious correlation. While it has developed scatter plots for all variables possibly influential in BP, are presented only to the three variables with higher correlation coefficients.

Table 4 - Descriptive statistics of quantitative variables

Studied variable	Minimum	Maximum	Mean	Standard Deviation	Variance
TP	18,1	28,3	25,8	2,1	4,5
UR	53,0	96,0	73,1	9,5	90,1
IP	0	296,5	96,5	80,4	6470,7
PA	53	397	140	56,2	3156,4
PE	0,006	34,524	3,946	4,667	21,780
TA	1,12	3,70	1,73	0,38	0,14
TE	146,410	237,100	157,721	24,201	585,697
SR	244,00	318,79	288,51	18,97	359,74
IF	-0,11	2,47	0,58	0,50	0,25
DO	2,21	3,59	2,81	0,33	0,11
TJ	15,79	26,32	19,45	3,48	12,09

N=3456

Figure 2 shows scatter diagrams: (a) P versus PE, (b) P versus TA and (c) PA versus D5. In Figure 2a there was a correlation imperfect, positively and with increased variability of PA as there is an increase of PE. Figures 2b and 2c, it was noted imperfect correlations, positive and with different PA profiles for each value of TA charged and socioeconomic class.

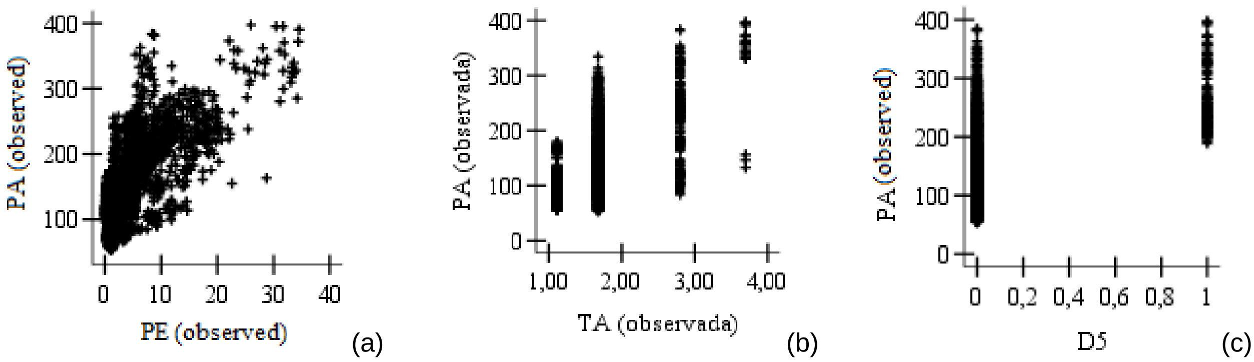


Figure 2 - Scatter diagrams

From: (i) correlation analysis, (ii) analysis of scatter diagrams, and (iii) the acceptance of the hypothesis that the residual analysis of a theoretical model any (mtQ) allows to obtain an appropriate model (mtA), as the result of (mtQ) added to his error (ε) is equal to the result of the (mtA), according to Equation 2, one can visualize a theoretical model any linear and additive, according to Equation 3.

$$mtQ + \varepsilon = mtA \quad (2)$$

$$mtQ = \beta_0+\beta_1.D2+\beta_2.D3+\beta_3.D4+\beta_4.D5+\beta_5.D7+\beta_6.D8+\beta_7.D9+\beta_8.D10+\beta_9.PE+\beta_{10}.TA \quad (3)$$

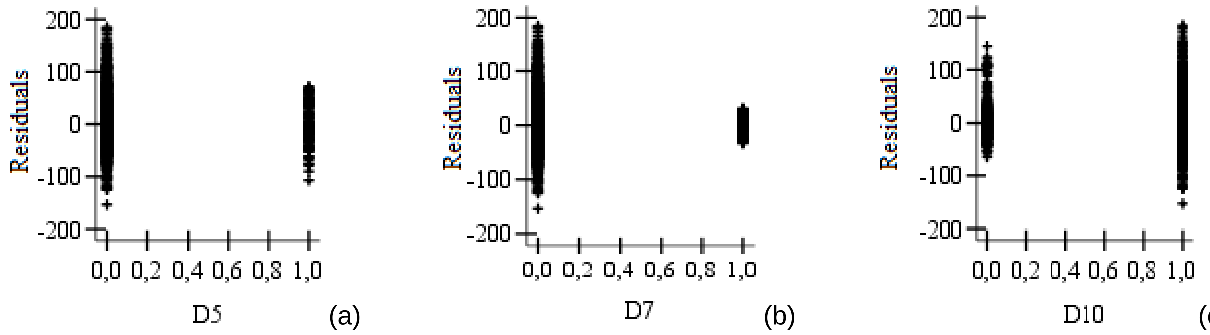
To adjust the model was necessary to remove the variable D9 ($p > 0.05$). In the evaluation and diagnosis met: $R^2 = 0.581$, F statistic ($p \leq 0.001$), presence of heteroscedasticity, non-normal distribution of residues and percentage error ranging from -120 to 60%, suggesting the need to adapt/change as expected. As a residual analysis was elaborated scatterplots of residuals versus explanatory variables (Figure 3) and hierarchical cluster analysis cluster (Figure 4).

The scatter plots revealed the existence of heteroscedasticity and residual profiles against different variables D5, D7 and D10. The hierarchical grouping cluster acknowledged the existence of two clusters, with the variables PE, D5, TA, D4 and D10 grouped the PA (Figure 4), reporting to homogeneity and association between these variables. These results led the proposition mtA divided by socio-economic strata, according to Equation 4.

Table 5 - Correlation coefficient for variables

	ME	TP	UR	IP	PA	D2	D3	D4	D5	D6	D7	D8	D9	D10	PE	TA	SR	IF	DO	TJ
ME	1	0,127**	-0,499**	-0,193**	0,033	0,000	0,000	0,000	0,000	-0,290**	-0,003	0,000	-0,013	0,014	0,025	0,002	0,484**	-0,433**	-0,402**	-0,296**
TP	0,127**	1	0,281**	0,566**	-0,005	0,000	0,000	0,000	0,000	0,532**	-0,009	-0,024	-0,010	0,023	0,023	-0,016	-0,280**	0,258**	0,034*	-0,184**
UR	-0,499**	0,281**	1	0,662**	-0,035*	0,000	0,000	0,000	0,000	0,647**	0,000	-0,031	0,010	0,006	0,004	-0,037*	-0,543**	0,419**	0,367**	0,115**
IP	-0,193**	0,566**	0,662**	1	-0,006	0,000	0,000	0,000	0,000	0,755**	-0,008	-0,035*	-0,007	0,026	0,029	-0,013	-0,535**	0,545**	0,248**	0,010
PA	0,033	-0,005	-0,035*	-0,006	1	-0,177**	0,051**	0,261**	0,439**	-0,008	-0,078**	-0,056**	-0,185**	0,241**	0,721**	0,526**	0,036*	-0,026	-0,043*	0,003
D2	0,000	0,000	0,000	0,000	-0,177**	1	-0,377**	-0,354**	-0,144**	0,000	0,017	0,039*	0,103**	-0,156**	-0,285**	-0,083**	0,000	0,000	0,000	0,000
D3	0,000	0,000	0,000	0,000	0,051**	-0,377**	1	-0,280**	-0,114**	0,000	-0,108**	-0,092**	-0,074**	0,127**	0,001	-0,030	0,000	0,000	0,000	0,000
D4	0,000	0,000	0,000	0,000	0,261**	-0,354**	-0,280**	1	-0,107**	0,000	-0,105**	-0,086**	-0,188**	0,271**	0,315**	0,188**	0,000	0,000	0,000	0,000
D5	0,000	0,000	0,000	0,000	0,439**	-0,144**	-0,114**	-0,107**	1	0,000	-0,043*	-0,035*	-0,076**	0,110**	0,535**	0,458**	0,000	0,000	0,000	0,000
D6	-0,290**	0,532**	0,647**	0,755**	-0,008	0,000	0,000	0,000	0,000	1	-0,015	-0,030	-0,001	0,022	0,030	-0,026	-0,510**	0,475**	0,162**	-0,007
D7	-0,003	-0,009	0,000	-0,008	-0,078**	0,017	-0,108**	-0,105**	-0,043*	-0,015	1	-0,034*	-0,075**	-0,387**	-0,147**	-0,173**	0,003	-0,002	0,006	-0,004
D8	0,000	-0,024	-0,031	-0,035*	-0,056**	0,039*	-0,092**	-0,086**	-0,035*	-0,030	-0,034*	1	-0,062**	-0,319**	-0,114**	-0,044*	-0,003	-0,011	0,022	-0,005
D9	-0,013	-0,010	0,010	-0,007	-0,185**	0,103**	-0,074**	-0,188**	-0,076**	-0,001	-0,075**	-0,062**	1	-0,695**	-0,233**	-0,146**	0,008	-0,001	-0,006	0,015
D10	0,014	0,023	0,006	0,026	0,241**	-0,156**	0,127**	0,271**	0,110**	0,022	-0,387**	-0,319**	-0,695**	1	0,361**	0,252**	-0,004	0,006	-0,011	-0,009
PE	0,025	0,023	0,004	0,029	0,721**	-0,285**	0,001	0,315**	0,535**	0,030	-0,147**	-0,114**	-0,233**	0,361**	1	0,495**	0,012	-0,007	-0,036*	-0,026
TA	0,002	-0,016	-0,037*	-0,013	0,526**	-0,083**	-0,030	0,188**	0,458**	-0,026	-0,173**	-0,044*	-0,146**	0,252**	0,495**	1	-0,005	0,036*	0,012	0,095**
SR	0,484**	-0,280**	-0,543**	-0,535**	0,036*	0,000	0,000	0,000	0,000	-0,510**	0,003	-0,003	0,008	-0,004	0,012	-0,005	1	-0,622**	-0,848**	-0,304**
IF	-0,433**	0,258**	0,419**	0,545**	-0,026	0,000	0,000	0,000	0,000	0,475**	-0,002	-0,011	-0,001	0,006	-0,007	0,036*	-0,622**	1	0,567**	0,440**
DO	-0,402**	0,034*	0,367**	0,248**	-0,043*	0,000	0,000	0,000	0,000	0,162**	0,006	0,022	-0,006	-0,011	-0,036*	0,012	-0,848**	0,567**	1	0,327**
TJ	-0,296**	-0,184**	0,115**	0,010	0,003	0,000	0,000	0,000	0,000	-0,007	-0,004	-0,005	0,015	-0,009	-0,026	0,095**	-0,304**	0,440**	0,327**	1

** Significant at $p < 0.01$
* Significant at $p < 0.05$



(a) (b) (c)

Figure 3 - Waste versus D5 (a) Residue versus D7 (b) versus residue and D10 (c)

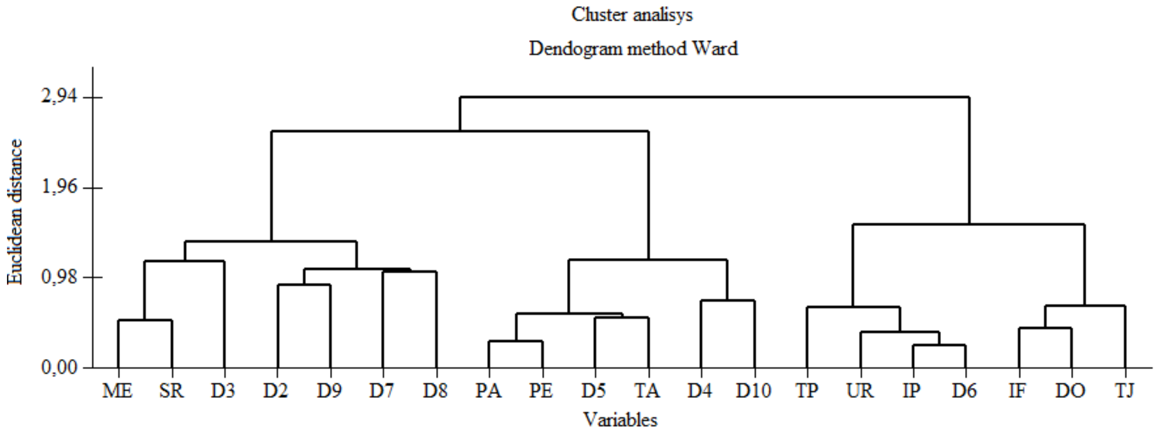


Figure 4 - Dendogram: Cluster Analysis

$$mtA = \begin{cases} mt_{E1} & para \quad D5 = 1 \quad D7 = 0 \quad D8 = 0 \quad D9 = 0 \quad D10 = 1 \\ mt_{E2} & para \quad D5 = 0 \quad D7 = 0 \quad D8 = 0 \quad D9 = 0 \quad D10 = 0 \\ mt_{E3} & para \quad D5 = 0 \quad D7 = 1 \quad D8 = 0 \quad D9 = 0 \quad D10 = 0 \\ mt_{E4} & para \quad D5 = 0 \quad D7 = 0 \quad D8 = 1 \quad D9 = 0 \quad D10 = 0 \\ mt_{E5} & para \quad D5 = 0 \quad D7 = 0 \quad D8 = 0 \quad D9 = 1 \quad D10 = 0 \\ mt_{E6} & para \quad D5 = 0 \quad D7 = 0 \quad D8 = 0 \quad D9 = 0 \quad D10 = 1 \end{cases} \tag{4}$$

Where: mt_{E1} theoretical model stratum = 1; mt_{E2} theoretical model stratum = 2; mt_{E3} theoretical model stratum = 3; ..., mt_{E6} = theoretical model stratum 6. What is proposed in this model is the consumption of water extracts from socio-economic, agreeing with the work of Katz-Gerro (2002) and Silva (2008), the results are given by Equations 5, 6, 7, 8 and 9, Tables 6 and Figure 4.

Table 6 - Theoretical model, assessment and diagnosis for the strata of consumption

Stratum 1: $mt_{E1} = \beta_0+\beta_1.TA+\beta_2.PE+\beta_3.PE^2+\beta_4.PE^3$ (5)											
R^2	R^2 adjusted	Analysis of coefficients				ANOVA					
		Coef	Valor	t	sig	FV	SQ	gl	QM	F	*p*
0,856	0,851	β_0	195,331	14,734	0,000	Reg	335659,96	4	83914,99	205,82	0,000
		β_1	37,203	-5,288	0,000	Res	56670,01	139	407,69		
		β_2	-14,669	12,443	0,000	Tot	392329,97	143			
		β_3	1,004	5,892	0,000						
		β_4	-0,016	-5,330	0,000						

The models mt_{E1}, mt_{E2}, mt_{E3}, and mt_{E4} mt_{E5} showed significant coefficients for the t test, highly significant F statistic, R^2 satisfactory ($0.617 \leq R^2 \leq 0.894$), residual fairly normal distribution (Figures 4a, 4d, 4g, 4j, 4n) reasonable homoscedasticity (Figures 4b, 4e, 4h, 4l, 4o) and residual mean percentage of ±30% (Figures 4c, 4f, 4i, 4m, 4p), which allowed the conclusion, with 1% risk, there regression between the PA and its explanatory variables. The values of R^2 can be compared to other studies of water consumption in cities, among them: Fields and Sperling (1996) – R^2 = 0.942, Oshima and Kosuda (1998) – R^2 = 0.985, Fernandes-Neto *et al*. (2004) – R^2 = 0.369; and Silva *et al*. (2008) – R^2 = 0.795.

Table 6 - Theoretical model, assessment and diagnosis for the strata of consumption (continued)

Stratum 2: $mtE2 = \beta0+\beta1.Ln(PE)+\beta2.(Ln(PE))2+\beta3.(Ln(PE))3+\beta4.(ln(PE))4+\beta5.TA-1$ (6)

R^2	R^2 adjusted	Analysis of coefficients				ANOVA					
		Coef	Valor	t	sig	FV	SQ	gl	QM	F	*p*
0,899	0,894	β_0	565,479	10,985	0,000	Reg	71176,81	5	14235,36	166,63	0,000
		β_1	415,891	4,841	0,000	Res	8884,81	104	85,43		
		β_2	206,855	4,143	0,000	Tot	80061,62	109			
		β_3	44,597	3,826	0,000						
		β_4	3,454	3,659	0,000						
		β_5	-211,22	-27,627	0,000						

Stratum 3: $mt_{E3} = \beta_0+\beta_1.D2+\beta_2.TA+\beta_3.PE+\beta_4.Ln(TA).PE^{-1}+\beta_5.Ln(TA).PE^{-2}+\beta_6.TA.PE^{-1}$ (7)

R^2	R^2 adjusted	Analysis of coefficients				ANOVA					
		Coef	Valor	t	sig	FV	SQ	gl	QM	F	*p*
0,634	0,617	β_0	241,081	6,091	0,000	Reg	58788,00	6	9798,01	37,85	0,000
		β_1	20,778	4,449	0,000	Res	33910,40	131	258,86		
		β_2	232,599	4,982	0,000	Tot	92698,41	137			
		β_3	-297,556	-6,161	0,000						
		β_4	-195,642	-3,705	0,000						
		β_5	84,823	5,894	0,000						
		β_6	-101,660	-7,676	0,000						

Stratum 4: $mt_{E4} = \beta_0+\beta_1.D3+\beta_2.PE+\beta_3.PE^2+\beta_4.PE^3+\beta_5.TA.PE$ (8)

R^2	R^2 adjusted	Analysis of coefficients				ANOVA					
		Coef	Valor	t	sig	FV	SQ	gl	QM	F	*p*
0,848	0,846	β_0	143,739	6,301	0,000	Reg	532286,92	5	106457,38	448,55	0,000
		β_1	9,349	3,722	0,000	Res	95646,08	403	237,33		
		β_2	-322,885	-5,558	0,000	Tot	627932,99	408			
		β_3	273,069	5,983	0,000						
		β_4	-67,612	-6,109	0,000						
		β_5	52,419	14,198	0,000						

Stratum 5: $mt_{E5} = \beta_0+\beta_1.PE.Ln(PE)+\beta_2.PE^2+\beta_3.TA^{-1,5}$ (9)

R^2	R^2 adjusted	Analysis of coefficients				ANOVA					
		Coef	Value	t	sig	FV	SQ	gl	QM	F	*p*
0,832	0,831	β_0	400,838	15,719	0,000	Reg	522240,91	3	174080,30	668,71	0,000
		β_1	616,921	11,321	0,000	Res	105430,19	405	260,32		
		β_2	-250,044	-9,927	0,000	Tot	627671,10	408			
		β_3	-73,516	-13,774	0,000						

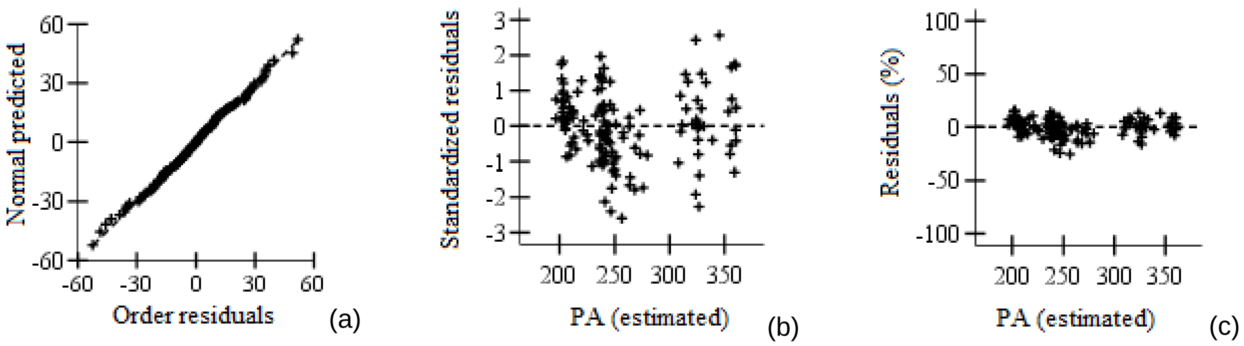


Figure 4 - Residual Analysis: mt_{E1} (a, b, c); mt_{E2} (d, e, f); mt_{E3} (g, h, i); mt_{E4} (j, l, m); mt_{E5} (n, o, p)

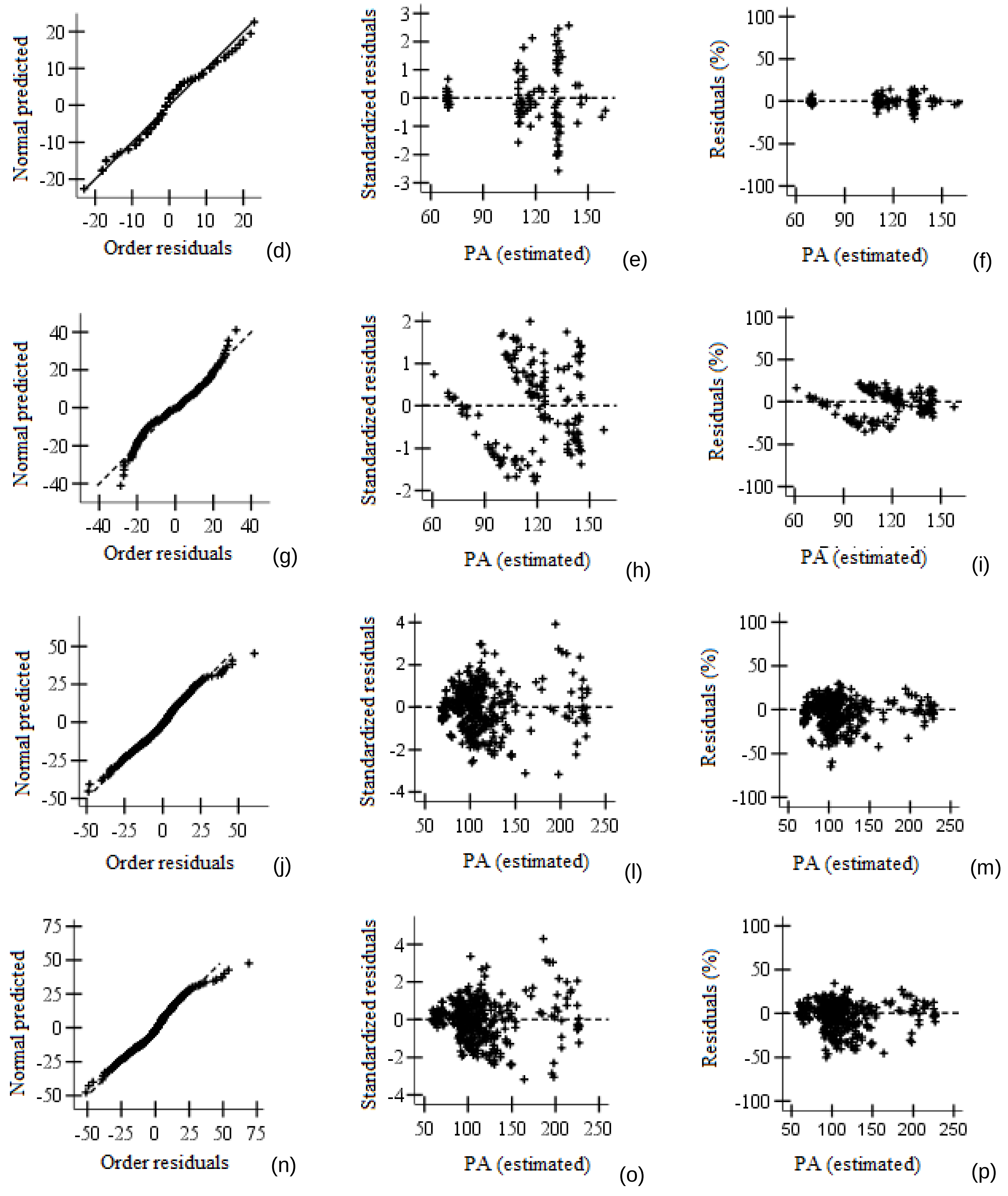


Figure 4 - Residual Analysis: mt_{E1} (a, b, c); mt_{E2} (d, e, f); mt_{E3} (g, h, i); mt_{E4} (j, l, m); mt_{E5} (n, o, p) (continued)

The search for mt_{E6} returned result present in the Equation 10, the data presented in Table 7 and Figure 5. The results showed significant coefficients for the t test, statistically highly significant F, R^2=0.585, problems in the test of normal distribution (Figure 5a), the presence of heteroskedasticity (Figure 5b) and percentage residue above 100% (Figure 5c). These observations make it very weak that any inferences regarding the consumption profile, just concluding that these variables are involved in the PA.

$$mt_{E6} = \beta_0+\beta_1.TA.PE +\beta_2.(TA.PE)^{0,5}.\ln(TA.PE)+\beta_3.(\ln(TA.PE))^2+\beta_4.(TA.PE\ )^{0,5}+\beta_5.\ln(TA.PE)+ \beta_6.(D2.TA)^{1,5}+\beta_7.(D2.TA)^2 \quad (10)$$

The test of consistency and aggregation, performed via graphical analysis of observed versus estimated values and heteroscedastic t-test are shown in Figure 6 and Table 7 respectively.

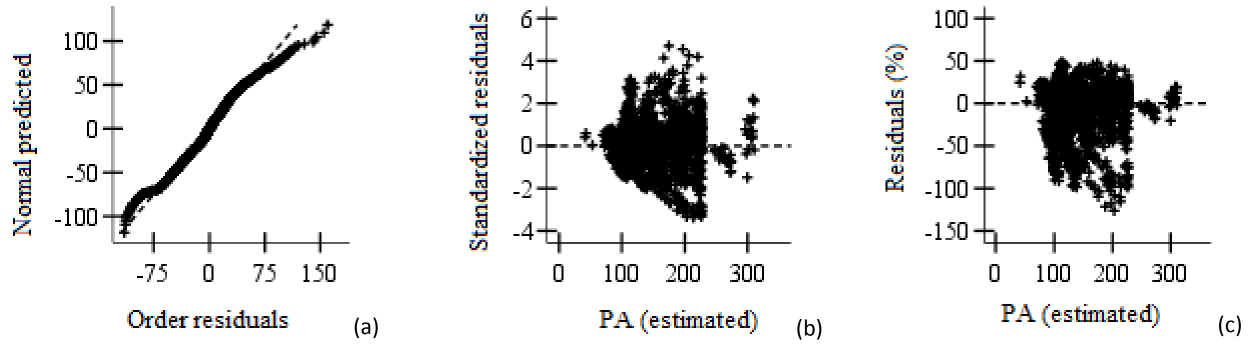


Figure 5 - Residual Analysis: mt_{E6}

In Figure 6, F1 represents the function of the ideal 1:1 line (observed values: the estimated values), allowing the conclusion that the models mt_{E1}, mt_{E2} mt_{E5} indicated reasonable care and the ideal function. For models mt_{E3}, mt_{E4} mt_{E6} and a trend towards non-compliance and linearity between the observed and predicted values, thus not reporting consistency and aggregation models. For mt_{E6} the result was expected since such issues had already submitted at the time of evaluation tests and diagnosis.

Table 7 - Stratum 6: mt_{E6} (Equation 10)

R^2	R^2 aj	Analysis of coefficients				ANOVA					
0,586	0,585	Coef	Value	t	sig	FV	SQ	gl	QM	F	*p*
		β_0	-13203,42	-2,989	0,002	Reg	4189710,60	7	598530,09	516,72	0,000
		β_1	134,660	2,338	0,019	Res	2956057,65	2552	1158,33		
		β_2	-2022,930	-2,867	0,004	Tot	7145768,25	2559			
		β_3	-806,644	-3,217	0,001						
		β_4	13112,450	3,013	0,002						
		β_5	-4531,638	-3,021	0,002						
		β_6	-57,095	-11,225	0,000						
		β_7	45,304	12,595	0,000						

Similar results were obtained by Karlis et al. (2007), when they studied the inclusion of the multiplicity and heterogeneity models to forecast the consumption of water in the city of Athens.

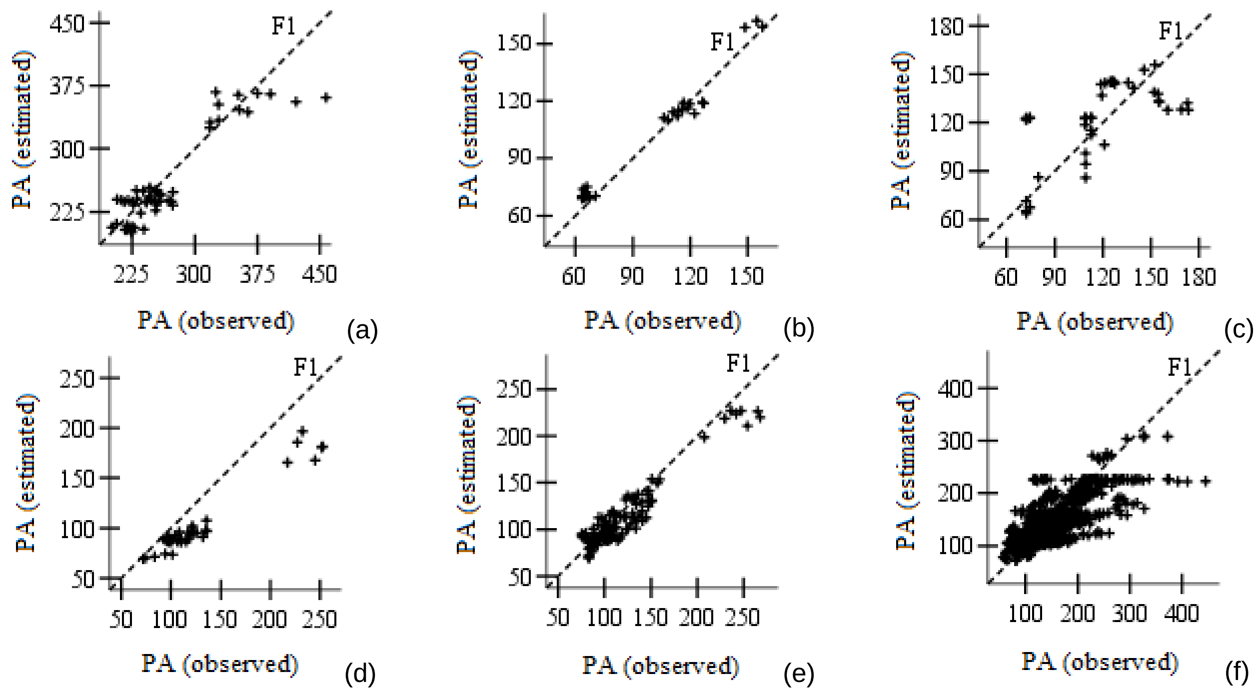


Figure 6 - Consistency and aggregation: (a) mt_{E1}, (b) mtE2, (c) mt_{E3}, (d) mt_{E4}, (e) mt_{P5}, (f) mt_{E6}

The heteroscedastic T-test, applied in order to test whether the mean observed and predicted values are significantly different or not, assuming they come from distributions with different variances (i.e., different samples from the same population have their results reported in Table 8).

Table 8 - Heteroscedastic Paired t-Test

Pair studied	Heteroscedastic paired T-test				
	V_{OP}	V_{mtP}	t	gl	sig
$PA_{OP1} - mt_{E1}$	3641,563	2976,285	0,634	93	0,528
$PA_{OP2} - mt_{E2}$	1033,640	915,769	-0,309	52	0,758
$PA_{OP3} - mt_{E3}$	910,251	583,148	-0,887	82	0,378
$PA_{OP4} - mt_{E4}$	2575,730	1205,460	2,469	65	0,016
$PA_{OP5} - mt_{E5}$	1598,967	1170,093	1,447	229	0,149
$PA_{OP6} - mt_{E6}$	3485,270	1774,644	2,984	1584	0,003

PAOP1, ..., = PA PAOP6 observed for the profile 1, ..., the profile observed for PA 6, VOP = variance observed for the profile; VMTP = estimated variance for the profile

The result of the paired t-test did not demonstrate any significant difference between the observed and estimated values for the models mt_{E1}, mt_{E2}, and mt_{E3} mt_{E5} ($p \leq 0.05$) and significant differences mt_{E4} and mtE6 ($p \leq 0.05$). Thus, the mt_{E1}, mt_{E2}, mt_{E3} mt_{E5} and can estimate the PA results significantly as close to real and otherwise have been the mtE4 and mtE6. As a possible explanation for the inadequacy of mt_{E4} there is the tendency to under-estimation of this model (Figure 6d), the inadequacy of mt_{E6} was expected due to problems stated above.

The correlation analysis, with the result shown in Table 9, showed strong associations (correlation coefficients ≥ 0.70) between the values observed for the stratum (PAOE) and the estimated values for the stratum (PAEE) for the models mtE1, mtE2, mtE4, and mtE5 mtE6. Since this criterion models can be considered as adequate.

Table 9 - Correlation coefficients for observed and estimated values

Studied pair	Correlation Coefficient
$PA_{OP1} - mt_{E1}$	0,913**
$PA_{OP2} - mt_{E2}$	0,988**
$PA_{OP3} - mt_{E3}$	0,625**
$PA_{OP4} - mt_{E4}$	0,973**
$PA_{OP5} - mt_{E5}$	0,951**
$PA_{OP6} - mt_{E6}$	0,749**

** Significant ($p < 0.01$)

PA_{OP1}, …, PA_{OP6} = PA observed for the profile 1, ..., PA observed for profile 6

Overall, there was acceptance of the hypothesis modeling of water consumption by extracts of socio-economic consumers. *mtA* presented component polynomial, exponential, inverse and interactions between variables, which exposes non-linearity and complexity in the phenomenon PA. The main variables affecting regional economic factors show how the key explanation of the PA, not being, in the case, significant climatic factor (TP, UR, IP) and macroeconomic factors (SR, IR, TJ, DO). We also emphasize the positive and imperfect relationship between TA and PA, ie the increase in the value of TA involving the BP increase.

One possible explanation is the price elasticity in the case of water demand inelastic (i.e., consumers are less sensitive to variations in price since there is a well that replace, according to Sullivan and Sheffrin, 2000). Also, the inappropriateness of estimating regression (ordinary least squares, OLS) in the context of prices would block another justification, according to Melo and Jorge-Neto (2007).These authors report that: (i) OLS ignores the possibility of grouping the observed data around the points of discontinuities, (ii) the progressive increase in the range of consumption makes the TA-dependent variable consumption (implying the violation of the independence hypothesis assumed by OLS), making the variable price of water an exogenous variable to the model, and (iii) an appropriate solution would be to use the model developed by Burtless and Hausman, who becomes non-linear budget constraints in the problem of utility maximization. Thus, reasonable fitness found by the mtPA can be attributed to modeling by clusters of variables (strata with different profiles of consumers) and the fact that the TA does not effectively represent the price of water, but the interaction between water tariff and the tracks consumption, which could explain the imperfect relationship between positive and TA and PA.

Conclusions

It was proposed a statistical model to predict the PA to the city of Cuiabá. The methodological procedures identified six socio-economic strata, three were adequate (mt_{E1}, and mt_{E2} mt_{E5}). The variables TS and D5 allowed the identification of consumer profiles (strata). The TA had a positive influence on BP, as a justification may be mentioned are the inelasticity of water, as Sullivan and Sheffrin (2000) and the interaction between the range of consumption and TA, as Melo and Jorge-Neto (2007).

Regarding the usefulness of the developed models one can mention the aid in water resources planning and orderly urban growth, since it allows to obtained estimates of future water demands, based on the typology of neighborhoods, and the estimated benefits from of possible measures for the conservation and rational use of water. It is suggested further studies: (i) alternatives such as cluster analysis combined with the model proposed by Burtless and Hausman (1978) or use of artificial intelligence techniques, which have shown good results in non-linear problems, (ii) check the influence of the precariousness of supply systems in water per capita consumption.

Acknowledgements

Thanks are due to company of sanitation of city Cuiabá (Sanecap), local energy company (Rede-Cemat) and National Institute of Meteorology (INMET) for their valuable support during data collection.

References

Aguirre, L.A. (2007). *Introdução a identificação de sistemas: Técnicas lineares e não-lineares aplicadas a sistemas reais*. Belo Horizonte, Federal University of Minas Gerais.

ANEEL - Agência Nacional de Energia Elétrica. (2002). Tarifas de fornecimento de energia elétrica. [on line]. available at: http://www.canalenergia.com.br/linktoaneel.asp?countn=ANEEL_Resolucoes__Externo&lnkRemoto=http://www.aneel.gov.br/cedoc/res2002180.pdf (accessed 26 November 2007).

Azevedo, C.S. (1992). Planejamento e Gerência no Enfoque Estratégico-Situacional de Carlos Matus. *Caderno de Saúde Pública*, **8**, 129-133.

BCB – Banco Central do Brasil. (2007). Cotação. Banco Central do Brasil. [on line]. available at: http://www5.bcb.gov.br/pec/taxas/port/ptaxnpesq.asp?id=txcotacao (accessed 08 June 2007).

BCB – Banco Central do Brasil. (2007). Histórico das taxas de juros. Banco Central do Brasil. [on line]. available at: http://www5.bcb.gov.br/pec/taxas/port/ptaxnpesq.asp?id=txcotacao (accessed 23 March 2007).

Campos, H.M.; Sperling, M. (1997). Proposição de Modelos para Determinação de Parâmetros de Projeto para Sistemas de Esgotos Sanitários com Base em Variáveis de Fácil Obtenção, in: proceedings of the Engenharia Sanitária e Ambiental Vol. 19, in Iguaçu, Brazil.

Carreira-Fernandez, J.; Garrido, R.J. (2002). *Economia dos recursos hídricos*. Salvador Edufba.

Conteiro, C.A.; Cantarutti, E.; Giovanni, S.C.F.; Morello, V. (2004). Premissas para realização de um Planejamento Estratégico. *Revista de Administração Nobel*, **3**, 77-86.

Cuiabá - Instituto de pesquisa e Desenvolvimento Urbano. (2006). *Perfil Socioeconômico de Cuiabá*. Cuiabá. Central de Texto.

Draper, N.R.; Smith, H. (1980). *Applied regression analysis*. New York. John Wiley and Sons.

Falkenberg, A. V. (2005). *Previsão de consumo urbano de água em curto prazo*. Masters dissertation, published by, PPMNE, Federal University of the Paraná, Curitiba, Brazil.

FEMA-MT. (2004). *Relatório Final Gerenciamento de Recursos Hídricos nas Vizinhanças da Cidade de Cuiabá/MT*. Cuiabá. Fundação Estadual de Meio Ambiente do Estado de Mato Grosso.

Fernandes-Neto, M.L.; Naghettini, M.; Von Sperling, M.; Libânio, M. (2004). Avaliação da relevância dos parâmetros intervenientes no consumo *per capita* de água para os municípios de Minas Gerais. *Revista ABES*, **9**(2), 100-107.

Gujarati, D. (2006). *Econometria Básica*. Rio de Janeiro. Elsevier.

IPEA – Instituto de Pesquisa e Economia Aplicada. (2007). Inflação - IGP-M. [on line]. available at: http://www.ipeadata.gov.br/ipeaweb.dll/ipeadata?322913109 (accessed 23 March 2007).

IPEA – Instituto de Pesquisa e Economia Aplicada. (2007). Salário mínimo real. [on line]. available at: http://www.ipeadata.gov.br/ipeaweb.dll/ipeadata?322913109 (accessed 23 March 2007).

IPECE - Instituto de Pesquisa e Estratégia Econômica do Ceará. (2003). *Estimativa de demanda de água residencial do Ceará*. Ceará. Governo do Estado do Ceará. Secretaria de Planejamento e Coordenação.

ISA – Instituto Socioambiental. (2007). *Abastecimento de água e esgotamento sanitário nas capitais brasileiras, em 2004*. [on line]. available at: http://www2.mananciais.org.br/upload_/abastesgotbrasilnov07.pdf (accessed 24 March 2008).

Karlis, D.; Vasdekis, V.G.S.; Banti, M. (2007). Heteroscedastic semiparametric models for domestic water consumption aggregated data. *Environmental and Ecological Statistics*.

Katz-Gerro, T. (2002). Highbrown cultural consumption and class distinction in Italy, Israel, West Germany, Sweden, and the United States. *Social Forces*, **81**, 207-229.

Lertpalangsunti, N.; Chan, C.W.; Mason, R.; Tontiwachwuthikul, P. (1999). A toolset for construction of hybrid intelligent forecasting systems: application for water demand prediction. *Artificial Intelligence in Engineering*, **13**, 21-42.

Lucion, C.E.R. (2005). Planejamento Financeiro. *Revista Eletrônica de Contabilidade*, **1**(3), 142-160.

Marroco, J. (2003). Análise estatística com utilização do SPSS. 2. ed. Lisboa. Silabo Ltda, 508 p.

Martinez-Espiñeira, R. (2002). Residential Water Demand in the Northwest of Spain. *Environmental and Resource Economics*, **21**, 161–187.

Melo, J.A.M.; Jorge-Neto, P.M. (2007). Residential Water Demand in the Northwest of Spain. *Pesquisa e Planejamento Econômico*, **37**(1), 149–173.

Mukhopadhyay, A.; Akber, A.; Al-Awadi, E. (2001). Analysis of freshwater consumption patterns in the private residences of Kuait. *Journal Urban Water*, **3**, 53–62.

Neter, J.; Wasserman, W. (1998). Applied linear statistical models: regression, analysis of variance and experimental designs. Georgia. Richard D. Irwin.

Oshima, N.; Kosuda, T. (1988). Distribution reservoir control with demand prediction using deterministic-chaos method. *Water Science and Techonology*, **37**(12), 389-395.

Piaia, I.I. (1997). Geografia do Mato Grosso. Cuiabá. EDUNIC.

Rede-Cemat – Centrais Elétricas Matogrossenses S.A. (2007). Consulta a Tarifas. [on line]. available at: http://www.gruporede.com.br/cemat/info_consultatarifa.asp (accessed 18 April 2007).

Rossetto, A.M.; Orth, D.M.; Rossetto, C.R; Flores, G.L. (2004). Proposta de um Sistema de Indicadores para Gestão de Cidades visando ao Desenvolvimento sustentável, in in: proceedings of the Cadastro Técnico Multifinalitário Vol. 6, in Florianópolis, Brazil.

Sanecap - Agência de Saneamento da Capital. (2007). Tarifa do consumidor. [on line]. available at: https://sanecap.locaweb.com.br/produtos.php?IDCategoria=306 (accessed 18 April 2007).

Setti, A.A.; Lima, J.E.F.W.; Chaves, A.G.M.; Pereira, I.C. (2001). Introdução ao gerenciamento de recursos hídricos. Brasília. Agencia Nacional de Energia Elétrica; Agência Nacional de Águas.

Silva, A.L.L. (2008). Consumo de produtos culturais em São Paulo: análise de fatores precedentes e proposta de modelo. PhD thesis, published by FEA-USP, São Paulo University, São Paulo, Brazil.

Silva, R.T.; Rocha, W.S. (1999). Caracterização da demanda urbana de água. Brasília. Secretaria Especial de Desenvolvimento Urbano.

Silva, W.T.P.; Silva, L.M.; Chichorro, J.F. (2008). Gestão de recursos hídricos: perspectivas do consumo *per capita* de água em Cuiabá. *Revista ABES*, **13**(1), 8-14.

SNIS – Sistema Nacional de Informações de Saneamento. (2004). Visão Geral da Prestação dos Serviços de Água e Esgotos – 2004. Sistema Nacional de Informações sobre Saneamento. Programa de Modernização do Setor de Saneamento. Ministério das Cidades.

Sokal, R.R.; Rohlf, F.J. (1995). Biometry: the principles and statistics in biological research. New York. W.H. Freeman and Company.

Souza, G.S. (1998). Introdução aos modelos de regressão linear e não-linear. Brasília. Empraba-SPI.

Sullivan, O.S.; Sheffrin, S.M. (2000). Princípios de Economia. Rio de Janeiro. LTC – Livros Técnicos e Científicos.

Zhou, S.L.; McMahon, T.A.; Walton, A.; Lewis. J. (2000). Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*. **236**. 153–164.

Zhou, S.L.; McMahon, T.A.; Walton, A.; Lewis. J. (2002). Forecasting operational demand for an urban water supply zone. *Journal of Hydrology*. **259**. 189–202.