Journal of Urban and Environmental Engineering, v.4, n.2, p.74-80

ISSN 1982-3932
doi: 10.4090/juee.2010.v4n2.074080

Journal of Urban and Environmental Engineering

www.journal-uee.org

DETERMINING INDICATORS OF URBAN HOUSEHOLD WATER CONSUMPTION THROUGH MULTIVARIATE STATISTICAL TECHNIQUE

Gledsneli M.L. Lins[1]*, Walter S. Cruz[2], Zedna M.C.L.Vieira[3], Francisco A.C. Neto[4] and Érico A.A. Miranda[5]

[1]*Natural Resources Graduation Program, Federal University of Campina Grande, Brazil*
[2]*Department of Civil Engineering, Federal University of Campina Grande, Brazil*
[3] *CAPES/National Postdoctoral Program Scholar, Federal University of Campina Grande, Brazil*
[4]*Civil Engineering Undergraduate Student, Federal University of Campina Grande, Brazil*
[5]*Department of Economy, Federal University of Campina Grande, Brazil*

Received 09 September 2010; received in revised form 30 November 2010; accepted 15 December 2010

Abstract: Water has a decisive influence on populations' life quality – specifically in areas like urban supply, drainage, and effluents treatment – due to its sound impact over public health. Water rational use constitutes the greatest challenge faced by water demand management, mainly with regard to urban household water consumption. This makes it important to develop researches to assist water managers and public policy-makers in planning and formulating water demand measures which may allow urban water rational use to be met. This work utilized the multivariate techniques Factor Analysis and Multiple Linear Regression Analysis – in order to determine the participation level of socioeconomic and climatic variables in monthly urban household consumption changes – applying them to two districts of Campina Grande city (State of Paraíba, Brazil). The districts were chosen based on socioeconomic criterion (income level) so as to evaluate their water consumer's behavior. A 9-year monthly data series (from year 2000 up to 2008) was utilized, comprising family income, water tariff, and quantity of household connections (economies) – as socioeconomic variables – and average temperature and precipitation, as climatic variables. For both the selected districts of Campina Grande city, the obtained results point out the variables "water tariff" and "family income" as indicators of these district's household consumption.

Keywords: Household water consumption, factor analysis, linear regression, urban water demand, water efficient use

---

* Correspondence to: Gledsneli M. de L. Lins, Tel.: +55 83 3310 1157; Fax: + 55 83 3310 1388
E-mail: glins16@yahoo.com.br

INTRODUCTION

The access to potable water is one of the most serious problems that urban centres are facing today all over the world. Albeit Brazil is privileged in terms of freshwater resources – it detains about 14% of the world's water availability – this richness is very unequally distributed among its regions: Amazonia (North region) concentrates 70%; Central-West region, 15%; South and Southeast regions, 6% each one; and Northeast region, just 3% of these resources. In relation to the main consumptive water uses, Brazilian water resources are shared by irrigation (63%), human supply (18%), and industry (5%) (ANA, 2007).

Whether these resources are utilized rationally, i.e., with economic efficiency, social equity and environmental sustainability, in the long run they could become a competitive advantage which would put Brazil among those countries high on the human development index (Lanna, 2008). Nevertheless, Brazil's population and economic growth occurred during the 20th century led to a predatory exploration of its natural resources in general, and, particularly, of its water resources (ANA, 2007).

Thus, along with the problems resulting from water scarcity in the Northeast region and water pollution in the Southeast region, Brazilian's large urban centres – like São Paulo e Recife – as well as medium size ones – like Campina Grande, State of Paraíba – have faced water supply crises which subjected population to tight rationing measures (Tomaz, 2001). This situation makes evident the need for integrated water resources management, and, more emphatically, the need for urban water demand management measures.

Water demand management (WDM) involves actions and sound methods to push the community in the direction of an appropriate use of water thus reducing water consumption by the final user and, at the same time, incrementing new consumption habits that would bring no impairment to comfort and hygienic necessities as provided by existing systems (USEPA, 1998). Then, water demand management goes beyond consumption management: rather than a question of organizing consumption data and of building graphics, WDM insists on studying these data and on guaranteeing the system feedback (Lins & Ribeiro, 2007).

The need for urban water demand management measures, which includes public policies for stimulating household water consumption efficiency, exceedingly justifies the development of researches that can indicate paths to address this objective.

Within this context, this work aims to analyze possible relationships between socioeconomic variables, climatic factors, and household water consumption for two districts of Campina Grande, Paraíba, Brazil, based on the multivariate statistical techniques Factor Analysis and Multiple Linear Regression Analysis and considering water potentiality and availability, so as to assist water managers and public policy-makers in planning and regulating water consumption.

STUDY CASE

Study area characterization

Campina Grande is the second largest city in the State of Paraíba, with a population of over 380 000 inhabitants (IBGE, 2010), which is divided along 49 districts. It is an important educational, industrial and technological centre situated in a semiarid region in northeastern Brazil. Located in the highest portion of the Borborema Plateau's oriental cliffs, with altitudes ranging from 500 to 600 m (average altitude of 551 m), the city presents a semiarid equatorial climate, average temperatures around 25°C, and an average annual rainfall of 730 mm (PMCG, 2002). Situated within the limits of the Paraíba River basin, Campina Grande, however, lies closely to the basin boundaries and, consequently, is not served by the main river. **Figure 1** shows the city's localization.

Fig.1 Maps: Brazil and Paraíba (Campina Grande city).

Water supply has become a historical problem. In the period from 1997 to 1999, the severe drought that fell upon the northeastern region of Brazil aggravated the critical water storage of the Epitácio Pessoa Reservoir that supplies the city and brought about a very serious crisis in water supply, affecting Campina Grande and other five cities in Borborema's region and subjecting 500 000 people to tight rationing measures (54% less water in its final phase) between November 1998 and April 2000 (Rêgo *et al.*, 2000; Galvão *et al.*, 2001).

In 2001 and 2002, during the dry season, new rationing measures had to be adopted motivated by the very low storage levels in the reservoir; in 2003, due to the same reason, a system of consumption quotas was adopted, for a few months, for all consumers; reverting this situation, a considerable increase inflow into the reservoir, registered in January 2004, allowed the accumulation of the reservoir's maxim storage volume (411 686 287 m^3) and avoided a new – and already planned – water rationing (Vieira *et al.*, 2005).

An analysis of the many factors that contributed to these crises – the city's geographic location, the lack of proper water resources management, the adoption of rationing policies only at times of impending crisis, among others – indicates the necessity for effective water resources management, particularly considering the demand side management and its potential for inducing rational consumption, in order to avoid the occurrence of new crises in the city's water supply (Vieira, 2002).

Nevertheless, it is important to emphasize that, in spite of these crises' seriousness and except for the rationing measures, no other water demand management initiatives have been taken in Campina Grande. Besides some lack of political will, the need for knowledge about household consumption characteristics and relationships – which could guide public policies formulation – appears as one of the factors that have made it difficult to implement urban water demand management in Campina Grande.

As an example of the very few existing works that address this city's necessity, Lins *et al*. (2008), utilized Factor Analysis and Simple Linear Regression Analysis (based on data relative to the year 2000) to identify the possible interrelations between the variable "population per age group" and "household water consumption" in Campina Grande city; the obtained results indicate that, for the predominant age group (0 to 49 years old), the city's average household water consumption (103 L/inhabitant·day) is inferior to the present Brazilian average consumption.

Based on socioeconomic criterion (family income level), two districts of Campina Grande city were chosen as study area:

- Dinamérica district, comprising an area of 1.33 km^2, a population of 3626 inhabitants (90.7% of which are included within the age group "0 to 49 years old") and an average family income of US$1,612, which characterizes it as a *medium income* district; and
- Mirante district, comprising an area of 0.52 km^2, a population of 1056 inhabitants (88% of which within the age group "0 to 49 years old") and an average family income of US$6,741, which characterizes it as a *high income* district (IBGE, 2008–2009 & SEPLAN/CG, 2000). **Figure 2** shows these districts' location in Campina Grande city's map.

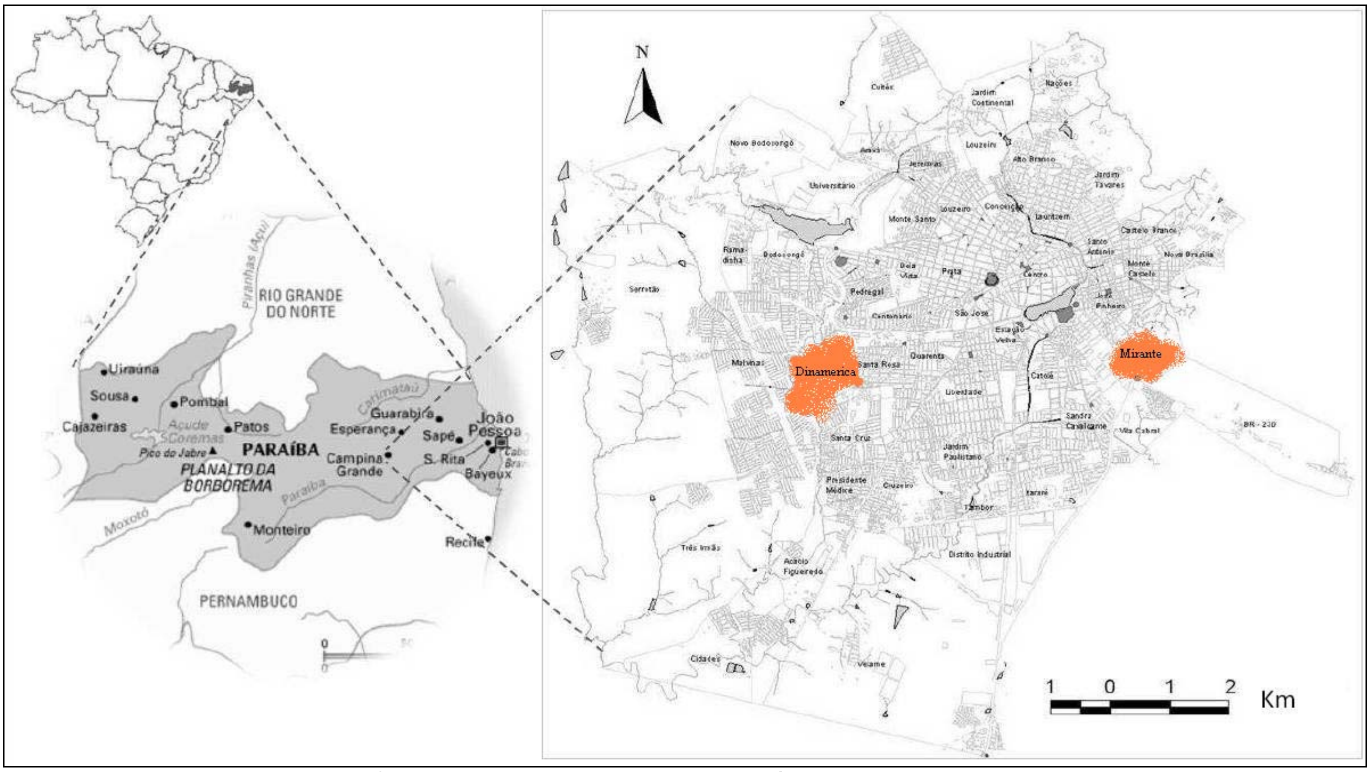

Fig.2 Map of Campina Grande city, indicating Dinamérica and Mirante districts location.

METHODOLOGY

Database

The following types of monthly data – relating to the selected districts – were used in this work, all of them referring to the period from 2000 to 2008 (9-year data series): *socioeconomic* (average family income, water tariffs applied by the local water supply company, and economies, i.e., the number of household connections to the public water supply system), *climatic* (average temperature and precipitation), and *household water consumption*. These data were made available by SEPLAN/CG (Campina Grande's Municipal Secretariat of Planning), EMBRAPA (Brazilian Enterprise for Agricultural Research), and CAGEPA (Water and Sewage Company of Paraíba, the local water supply company).

Statistical Techniques

This work utilized two multivariate statistical techniques – Factor Analysis and Multiple Linear Regression Analysis – which were made viable by the Statistical Package for the Social Sciences – SPSS.

At first, a Factor Analysis (FA) was carried out considering as variables, for both Dinamérica and Mirante districts: monthly average temperature, monthly average precipitation, economies, monthly water tariffs, and monthly family income.

Factor Analysis is a multivariate statistical analysis technique which changes the original variables set into a smaller one composed by factors; the order in which these factors are obtained corresponds to their importance with respect to the quantity of variance of the original variables set that each of them can explain. This means that the first factor explains the greatest quantity of variance of the data set; the second factor explains the greatest possible quantity of remainder variance; and so on. As this work deals with a 5-variable set, the maximum number of factors to be extracted is five, i.e., equal to the number of original variables. Here, the factoring (extraction of factors) method utilized was Principal Component Analysis (PCA).

On the other hand, as the quantity of variance explained by each factor is represented by the quotient between the factor's eigenvalue and the number of variables (or the total number of factors), it is considered, in general, that only those factors with eigenvalues greater than or equal to 1 will take part in the final solution. In other words, factors' eigenvalues which are less than 1 are eliminated from the final solution (Cruz, 1983).

Generally, the factors thus extracted are not easily interpreted. In order to facilitate their interpretation, it is realized a rotation of the coordinate axes which graphically represent the factors. According to the rotation method adopted – orthogonal or oblique – the variables' coordinates, also called "factors loading," represent the projection of the variables' variance in this new coordinate system. One of the most utilized methods is the so-called Varimax rotation which aims to maximize the variance projection of each variable on the factors. This was the rotation method used in this work.

The other multivariate analysis technique utilized was Multiple Linear Regression Analysis. This technique shows the linear combination between a variable, called *dependent variable*, and a set of variables called *independent* or *explicative variables* (Matos, 2000).

This work adopted the significance level of 5% and, for both districts, considered the "household consumption" (calculated, for each district, as the quotient between the total district water consumption and the number of economies) as dependent variable; the variables initially suggested as independent variables were those which have presented the greatest loadings on Factor 1, the most important factor. Based on this set of variables thus defined, a Multiple Linear Regression Analysis was performed, and, through the "stepwise method" (i.e., a method executed step-by-step, which includes each independent variable according to its contribution for increasing the determination coefficient of the linear regression equation), the final linear regression equation was obtained.

RESULTS FOR DINAMÉRICA DISTRICT

Factor Analysis Application

Table 1 shows the first obtained results for Dinamérica district, indicating the initial eigenvalues and the percentages of variance for each one of the five factors.

Based on the criterion of eigenvalues greater than 1, two factors were extracted (Factor 1 and Factor 2) which explain nearly 76% of the original data set variance.

Following the methodological steps described above, the Varimax method application resulted as shown in **Table 2**. In this case, the socioeconomics variables (Tariff, Economies and Income) represent a high loading on Factor 1, while the climatic variables (average temperature T, and average precipitation *P*) represent a high loading on Factor 2.

Table 1. Total variance explained

Factor	Eigenvalue	Variance (%)	Cumulative (%)
1	2.473	49.469	49.469
2	1.338	26.768	76.238
3	0.655	13.109	89.347
4	0.481	9.629	98.976
5	$5.12.10^{-2}$	1.024	100.000

Extraction Method: Principal Component Analysis.

Table 2. Rotated matrix

Variables	Factor 1	Factor 2
T	6.882×10^{-2}	0.831
P	-2.458×10^{-2}	- 0.825
Tariff	0.975	9.846×10^{-2}
Economies	0.860	6.362×10^{-2}
Income	0.855	-3.907×10^{-2}

Rotation Method: Varimax rotation.

Multiple Linear Regression Analysis Application

The Multiple Linear Regression Analysis between "household consumption" and the explicative variables extracted from the Factor Analysis (Factor 1) presented the final regression equation as expressed by **Eq. (1)**:

$$Y = 1.933 + 0.131\, X_2 - 8.246\times10^{-4}\, X_3 \qquad (1)$$

where Y is the household consumption (m^3/month); X_2 is the water supply company's tariff; and X_3 is the monthly average family income.

This final regression equation's determination coefficient R^2 is equal to 0.896, which means that 89.6% of the dependent variable Y's variance are explained by the independent variables X_2 (Tariff) and X_3 (Income).

Table 3 shows the Analysis of Variance (ANOVA) for **Eq.(1)**.

Based on **Table 3** analysis, one can verify that statistics F presents a value equal to 447. Admitting a given level of significance, this F's value must be compared to a value obtained from a "Table of F's values," which considers the regression and residual number of degrees of freedom (*df*). Only if the F's value presented in **Table 3** is greater than any F's values tabulated, for the considered significance level and degrees of freedom, it will be possible to say that **Eq.(1)** is *statistically significant* at that adopted significance level.

It is important to emphasize that SPSS' ANOVA straightly indicates the value of the smallest significance level (the sixth column of **Table 3**) from which **Eq. (1)** is statistically significant. In other words, **Table 3** (at its fifth column) indicates the F's value which is greater than any tabulated F for significance levels greater than that one presented at its sixth column.

Since the adopted significance level (5%, i.e., 0.05) is greater than the smallest significance level (0.00) showed by **Table 3**, it is possible to affirm that $F = 447$

Table 3 – ANOVA

Model	Sum of squares	*df*	Mean square	*F*	Sig
Regression	21,81	2	10,9	447	.00
Residual	2,53	104	,024		
Total	24,34	106			

df = degrees of freedom; *F* = Fisher's statistics;
Sig = Significance level.

Table 4. Dinamérica district: coefficients of linear regression equation model

Model	*B*	*t*	Sig
Constant	1.933	39.746	.000
Tariff	.131	27.699	.000
Income	-8.246×10^{-4}	-15.579	.000

B = Unstandardized coefficient; *t* = t-Student;
Sig = Significance level.

is greater than the tabulated value of F for the 5% – significance level, and, consequently, the obtained final regression equation is *statistically significant* at this adopted significance level.

Table 4 indicates the constant part, the partial regression coefficients, as well as other information which are pertinent to the statistical quality analysis of the independent variables in **Eq. (1)**.

As **Table 4** shows, the values assumed by statistics t are greater than those tabulated for any significance level which is greater than 0.000 (result presented by SPSS at the fourth column of **Table 4**). Thus, the partial regression coefficients are statistically significant at the adopted significance level (5%).

Therefore, for Dinamérica district, **Eq. (1)** can be considered as a linear regression equation model which associates household consumption to the independent variables Tariff and Income.

RESULTS FOR MIRANTE DISTRICT

Factor Analysis Application

Following the methodological steps, **Table 5** shows the first results for Mirante district, i.e., the initial eigenvalues and the percentages of variance for each one of the five factors. Again, the criterion of eigenvalues greater than 1 implied in the extraction of two factors, which explain about 79% of the original data set variance.

The rotated matrix (Varimax Method), shown by **Table 6**, presents a result which is similar to that obtained for Dinamérica district, i.e., the socioeconomic variables (Tariff, Economies and Income) presented a high loading on Factor 1, while the same occurred with the climatic variable (T and P) in relation to Factor 2.

Table 5. Total variance explained

Factor	Eigenvalue	Variance (%)	Cumulative (%)
1	2.614	52.285	52.285
2	1.336	26.717	79.002
3	0.640	12.800	91.802
4	0.325	6.491	98.293
5	8.53×10^{-2}	1.707	100.000

Extraction Method: Principal Component Analysis.

Table 6. Rotated matrix.

Variables	Factor 1	Factor 2
T	5.491×10^{-2}	0.836
P	-5.394×10^{-2}	- 0.817
Tariff	0.941	0.175
Economies	0.944	4.598×10^{-2}
Income	0.876	-1.992×10^{-2}

Rotation Method: Varimax rotation.

Multiple Linear Regression Analysis Application

The linear regression analysis was performed by considering the district's household consumption and its possible relations to the explicative variables extracted from the Factor Analysis (Factor 1). The following final regression equation was obtained:

$$Y = 5.970 + 0.104X_2 - 6.13\times10^{-4}X_3 \quad (2)$$

where Y is the household consumption (m^3/month); X_2 is the water supply company's tariff; and X_3 is the monthly average family income.

Equation (2) has a determination coefficient R^2 equal to 0.912, which means that 91.2% of the dependent variable Y's variance are explained by the independent variables X_2 (Tariff) and X_3 (Income).

Table 7 presents the results of the Analysis of Variance (ANOVA).

The value of statistics F is equal to 537, which is greater than any F's values tabulated for the 5% – significance level (in accordance with the explanation already given for **Table 3** analysis). Consequently, this means that **Eq. (2)** is *statistically significant* at this adopted significance level.

Table 8 shows the constant part, the partial regression coefficients, and other information related to the statistical quality analysis of the independent variables in **Eq. (2)**.

Table 7. ANOVA.

Model	Sum of squares	*df*	Mean square	*F*	Sig.
Regression	117.35	2	58.68	537	.000
Residual	11.36	104	.109		
Total	128.72	106			

df = degrees of freedom; *F* = Fisher's statistics;
Sig = Significance level.

Table 8. Mirante district: coefficients of linear regression equation model.

Model	*B*	*t*	Sig
Constant	5.970	55.059	.000
Tariff	.104	32.443	.000
Income	-6.134×10^{-4}	-26.716	.000

B = Unstandardized coefficient; *t* = t-Student;
Sig = Significance level.

As statistics t values are greater than those tabulated for any significance level greater than 0.000, the partial regression coefficients are statistically significant.

Therefore, for Mirante district, **Eq. (2)** can be considered as a linear regression equation model which associates household consumption to the variables Tariff and Income.

CONCLUSION

Household water consumption is a complex function of several factors, amongst which are socioeconomic and climate ones. For the study case considered in this work, and specifically for the two selected districts, it was verified that the variables "average temperature" and "average precipitation" did not take part in the final solution, due to the fact that none of them have presented a high loading on Factor 1, the most important factor that has been indicated by Factor Analysis application.

Each obtained linear regression equation, **Eq. (1)** and **Eq. (2)**, respectively for Dinamérica and Mirante districts, can be considered as a linear regression model which associates household consumption to water tariffs and family income.

Then, the obtained results point out the socioeconomics variables (water tariffs and family income) as indicators of urban household consumption for these Campina Grande city's districts, so that they can be utilized to assist decision-makers and to afford a more effective management of this consumption.

The authors suggest that other investigations must be realized, including different districts and/or sectors of Campina Grande city, so as to point out linear regression models which can incorporate any variability with regard to this city's household water consumption, as well as to support urban water demand public policies formulation.

Acknowledgements The third author gratefully acknowledges the scholarship provided by the Brazilian Agency CAPES, through the National Postdoctoral Program (PNPD).

REFERENCES

ANA – Brazilian National Water Agency (2007) *Relatório nacional sobre o gerenciamento da água no Brasil- GEO Brasil Recursos Hídricos* Available in: http://www.ana.gov.br. Accessed in: 15 May 2010.

Cruz, W.S. (1983) Métodos estatísticos multivariados aplicados à análise de transportes. MSc Thesis, Universidade Federal da Paraíba, Campus II, Campina Grande, Paraíba, Brasil.

Galvão, C.O., Rêgo, J.C., Ribeiro, M.M.R. & Albuquerque, J.P.T. (2001) Sustainability characterization and modeling of water supply management practices. *IAHS Publ.* 268, 81–88.

IBGE – Instituto Brasileiro de Geografia e Estatística (2010). Dados da cidade de Campina Grande – PB. Available in: http://www.ibge.org.br/cidades. Accessed in: 20 May 2010.

IBGE – Instituto Brasileiro de Geografia e Estatística (2008-2009) Pesquisa de orçamentos domiciliares. Available in: http://www.ibge.org.br/estadosat/. Accessed in: 20 May 2010.

Lanna, A.E.L. (2008) A economia dos recursos hídricos: os desafios da alocação eficiente de um recurso (cada vez mais) escasso. *Estudos Avançados*, **22(2)**, 113–130. doi: 10.1590/S0103-40142008000200008.

Lins, G.M.L. & Ribeiro, M.M.R. (2007) Gestão da demanda de água em centros urbanos do semiárido nordestino. *Proc. XVII Seminário Brasileiro de Recursos Hídricos*, ABRH, São Paulo, SP, Brasil.

Lins, G.M.L., Cruz, W.S. & Miranda, E.A.A (2008) O uso de técnicas estatísticas multivariadas na determinação de indicadores de consumo doméstico urbano de água. *Proc. IX Simpósio de Recursos Hídricos do Nordeste*, ABRH, Salvador, Brazil, 25–28 Nov.

Matos, O.C. (2000) *Econometria Básica: teoria e aplicações*. São Paulo: Atlas.

PMCG – Prefeitura Municipal de Campina Grande (2002) Perfil do município de Campina Grande. Available in: http://www.pmcg.pb.gov.br. Accessed in: 27 May 2010.

Rêgo, J.C., Ribeiro, M.M.R. & Albuquerque, J.P.T. (2000) Uma análise da crise de 1998–2000 no abastecimento d'água de Campina Grande-PB. *Proc. V Simpósio de Recursos Hídricos do Nordeste*, ABRH, Natal, RN, Brasil.

SEPLAN/CG – Secretaria de Planejamento de Campina Grande (2000) Dados dos bairros – geral/ In: *Perfil do município de Campina Grande*. Available in: http://www.pmcg.pb.gov.br. Accessed in: 27 May, 2010.

Tomaz, P. (2001) *Economia de água para empresas e residência: Um estudo atualizado sobre o uso racional da água.* São Paulo: Editora Navegar, 112p.

USEPA – U.S. Environmental Protection Agency (1998) Water Conservation Plan Guidelines. Available in: http://www.epa.gov/own. Accessed in: 12 December 2000.

Vieira, Z.M.C.L. (2002) Análise de conflitos na seleção de alternativas de gerenciamento da demanda urbana de água. MSc Thesis, Federal University of Campina Grande, 132p.

Vieira, Z.M.C.L., Braga, C.F.C. & Ribeiro, M.M.R. (2005) Conflict analysis as a decision support tool in urban water demand management. *IAHS Publ.* **293**, 65–72.