Water Utility Journal 25: 15-21, 2020.

Identification of correlation between residential water demand and average income using the pool regression model: Study case in Fortaleza - Brazil

T.A. Oliveira*, F.A. Souza Filho, G.A. Reis and T.M.N. Carvalho
Department of Hydraulic and Environmental Eng., Federal University of Ceará, Fortaleza, Brazil
* *e-mail: thais.antero@gmail.com*

Abstract: The increase in water consumption was twice as large as the increase in population during the last decades Thus, the uncertainty of water supply capacity is a greater need for efficient water resources management. For this to occur, the decision-making process must have a technical and scientific basis. To assist this process, this study aims to identify any statistical correlation between the various patterns of water consumption with the average income of the population in urban centers. The analysis is done by census tract, using data from the city of Fortaleza (CE) from the application of the multilevel linear regression model. To develop the analysis, monthly data for each census tract 2009 to 2017 were used, according to the IBGE boundaries. At the end of the statistical process, three equations are developed, which relate the intercept coefficients of the equations and the variation in time coefficients with income. In conclusion, the most substantial relationship is found between the average income and the initial value of demand (intercept of the equation). However, it does not present strong dependence on the variation coefficient of the residential water demand, which shows that it is impossible to estimate a relation of demand with the time of each sector from the knowledge of the average sector income.

Key words: water demand; urban supply; pooled data; SIG

1. INTRODUCTION

Facing increasingly frequent water crises, human supply became an even more complex challenge. In large urban centers, there are high residential water demand values due to changes in consumption patterns. Decision-makers are constantly finding themselves facing critical scenarios and yet it is essential to achieve a balance in supply and demand (Fielding et al., 2013).

During the last decades, the increase in water consumption was twice as significant as the increase in population. For the next few years, studies estimate that the increase in water demand, combined with water scarcity, contributes to the expansion of water stress in several regions of the world (Suarez-Alminana et al., 2017; Cosgrove and Loucks, 2015). Due to changes in nature and gaps in the observations, the uncertainty of water supply capacity requires more efficient management and more equitable use of water resources (Pienaar and Hughes, 2017). The process of decision-making, when developed from a robust technical and scientific basis, promoting benefits and efficiencies, will result in positive impacts on all levels: social, economic and environmental (Lianqing et al., 2012.).

Schleich and Hillenbrand (2008) affirmed that the per capita income may be a major factor to explain the discrepancy in per capita water demand. In order to provide inputs to the decision-making process, this work proposes to analyse the variation of demand in urban centers, to identify any statistical correlation between the variation patterns of water consumption and the average income. This article focuses on finding out if there is a relevant level of correlation between the variation of the water consumption and the income, in order to help understanding the spatial distribution differences in water demand variation. Therefore, to assist the water demand forecast process, which can promote more appropriate demand management policies (such as water meter installation, price changes, and the non-price rationing regulations in water use) implementation.

The analysis encompasses census sectors, using data from the city of Fortaleza (CE), Brazil, and applies the pool regression model. The city of Fortaleza is the state capital of Ceará, located in the northeastern Brazil. The city has, approximately, 2.5 million inhabitants and it is the fifth largest city in Brazil, with an area of 313 km², being one of the highest demographic densities in the country (8,001 per km²) (de Maria André and Carvalho, 2014). Ceará has more than 90% of its territory located in the Brazilian semiarid region, which has low precipitation levels, less than 800 mm per year, high evaporation rates, and shallow soils. These characteristics make the region strongly vulnerable to droughts. Regarding water consumption in Fortaelza, André (2012) reported that one of the determining factors to explain residential water consumption in the city is the consumers' income.

This model has been applied to clinical trial analysis (Rothwell et al., 2003; Stephens et al., 2018; Ward and Leigh, 1993), to political science studies (Li and Schaub, 2004; Saideman et al., 2002) and to water resources management works, more specifically, to residential water demand analysis (Danielson, 1979; Hanke and Maré, 1982).

2. METHODOLOGY

Outcome variables analyses are commonly continuous variables and, most fundamentally, are obtained by regularly measuring the same individuals over time (Ward and Leigh, 1993). For this study, the outcome variable is the residential water demand and the individuals are the census sectors. According to Ward and Leigh (1993), outcome evaluation can be arranged in a two-dimensional data matrix of n rows and t columns, each row indicating the results for one individual and each column representing the results at one-time point. This simple matrix is all that is necessary to examine the outcome variable time course. Nonetheless, in several cases, a three-dimensional data matrix is necessary, containing data on some additional variables also obtained for each individual at each time point. These additional variables can vary over time or not (time-invariant).

For this study, the additional variable is the average income for each census sector. Since the only source of the socioeconomic data for the sectors is the IBGE (Brazilian Institute of Geography and Statistics) census that is realized once in ten years, the variable is time-invariant. Finally, Ward and Leigh (1993) claim that this variable may be applied as an independent variable in pooled time series regression models to determine their relationships with, or effects on, the outcome.

The analysis used monthly water demand data (in m³) of each census sector from 2009 until 2017. Pooled time-series regression models do not have specific requirements for the time series length, the observations frequency, or the sample size. The study census sector spatial division that was adopted is the same as the 2010 IBGE census. Initially, in order to help visualize and understand the spatial distribution amongst the census sectors of Fortaleza, the authors developed two thematic maps: one to illustrate the average income and the other to represent the HDI.

To calculate the correlation between water consumption and average income, monthly data were grouped in annual sums, along with unit conversion (m³/month to L/day). The ratio between each treated demand and the corresponding sector population was calculated to get the per capita demand. Due to noise in the acquired data, the authors had to eliminate sectors that presented zero population.

The first regression analysis was performed between the annual demand data and time values, organized as shown in Table 1. The demand data is represented as $D_{i,t}$, in which "i" is an index indicating the sector and "t" is the corresponding year. The years were indicated with integers, starting at 0 (2009) to 8 (2017), to simplify data representation.

A second-order linear regression is developed through the R statistical analysis software to obtain a regression equation between demand and year of each sector, as in Equation 1.

$$D_{i,t} = \beta_i^1 + \beta_i^2 t + \beta_i^3 t^2 \quad (1)$$

So, three coefficients for each sector are obtained and it is possible to analyze the correlation of each with the average income data sectors. Thus, three equations were developed which relate by linear regression each coefficient with the average income. These equations are as follows:

$$\beta_i^1 = \alpha R_i + \alpha_0 \quad (2)$$

$$\beta_i^2 = \gamma R_i + \gamma_0 \quad (3)$$

$$\beta_i^3 = \delta R_i + \delta_0 \quad (4)$$

Table 1. Table model that relates each sector to the demands of each year

Sector Code	2009 demand (T = 0)	2010 demand (T = 1)	2011 demand (T = 2)	2012 demand (T = 3)	2013 demand (T = 4)	2014 demand (T = 5)	2015 demand (T = 6)	2016 demand (T = 7)	2017 demand (T = 8)
Sector 1	$D_{1,0}$	$D_{1,1}$	$D_{1,2}$	$D_{1,3}$	$D_{1,4}$	$D_{1,5}$	$D_{1,6}$	$D_{1,7}$	$D_{1,8}$
Sector 2	$D_{2,0}$	$D_{2,1}$	$D_{2,2}$	$D_{2,3}$	$D_{2,4}$	$D_{2,5}$	$D_{2,6}$	$D_{2,7}$	$D_{2,8}$
...	...	...	...	...	...	...	...	...	...
No sector	$D_{n,0}$	$D_{n,1}$	$D_{n,2}$	$D_{n,3}$	$D_{n,4}$	$D_{n,5}$	$D_{n,6}$	$D_{n,7}$	$D_{n,8}$

3. RESULTS AND DISCUSSION

The thematic maps (Figures 1 and 2) present that the income and distribution in Fortaleza are heterogeneous, with high levels of concentration in the northern part of the city. This area encompasses known luxurious neighborhoods, such as Aldeota and Meireles. Since the 1980s, these two neighborhoods became part of the main city centrality, concentrating most of the wealth and socially produced values, employment and income, capital and information flows, and political and business decision-making centers in Fortaleza (Machado, 2017).

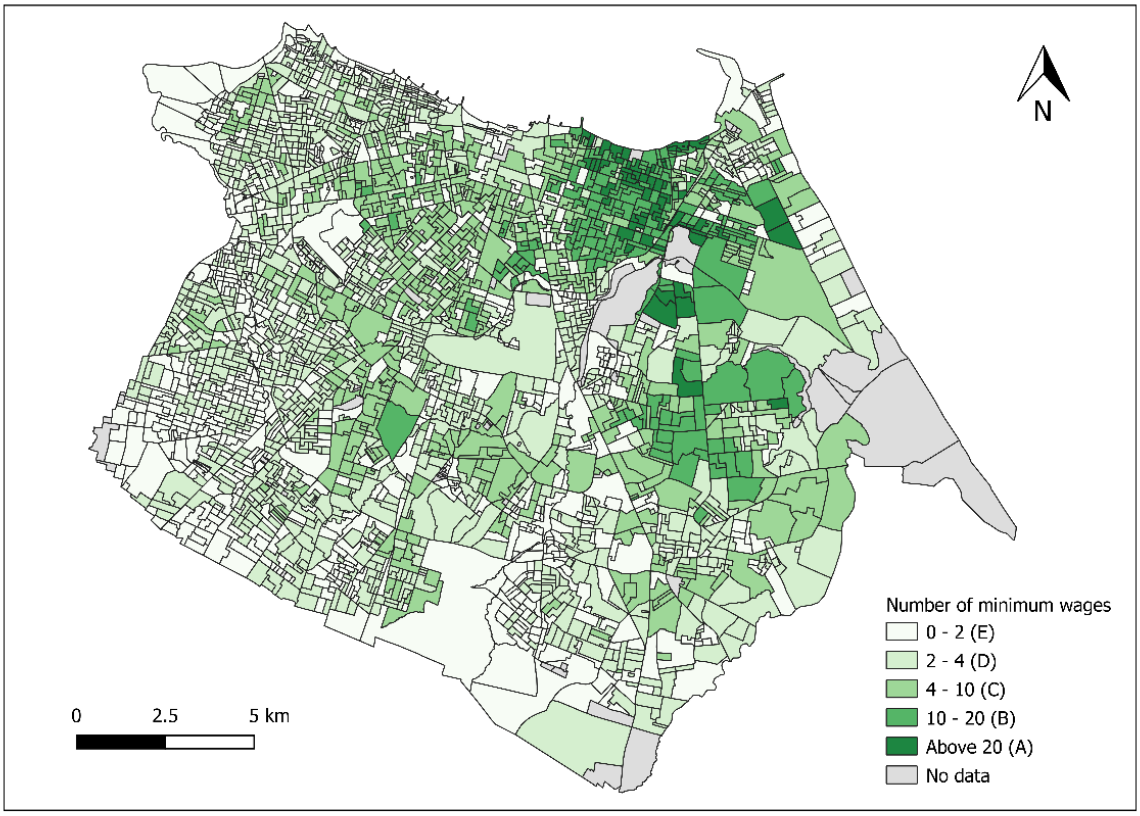

Figure 1. Classes of middle income of census sectors

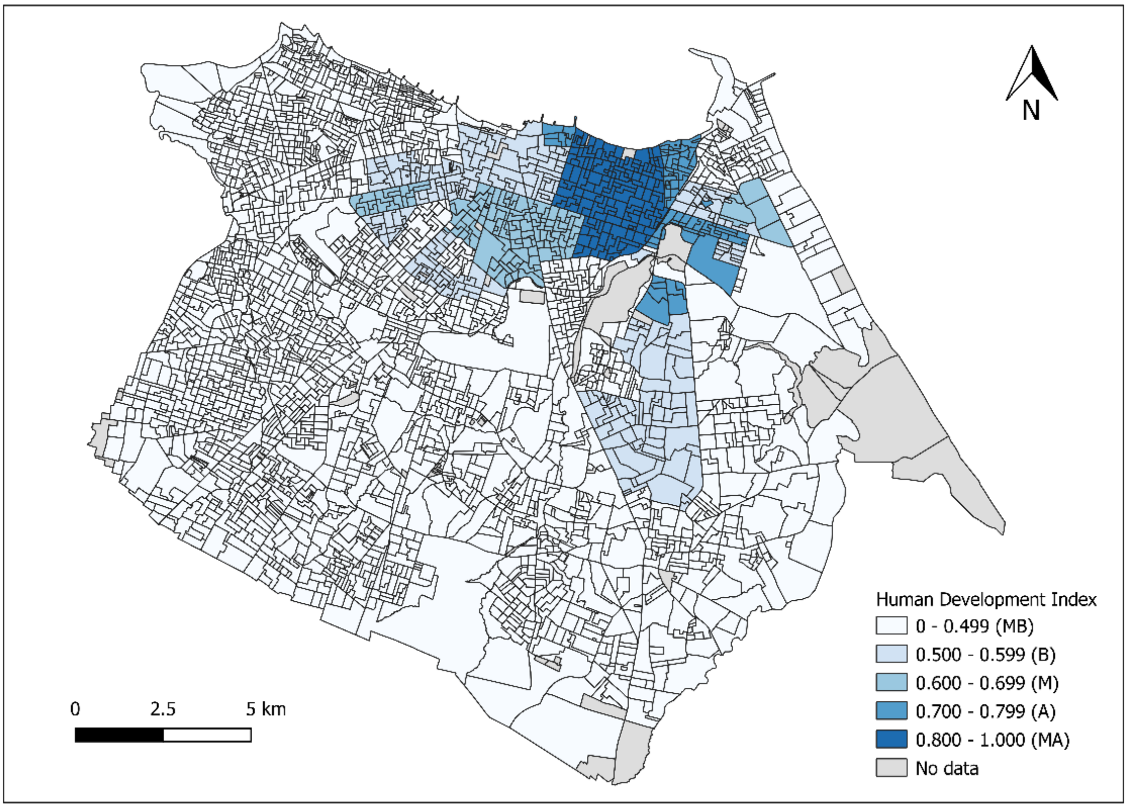

Figure 2. HDI classes of census sectors

Water demand data were grouped in order to organize a vast array with 2993 sectors and their values per capita. Due to the eliminated sectors with zero population, the array was reduced to 2952 sectors considered with appropriate data. After this processing, a second-order regression was performed to get the three coefficients. For instance, the sector with the number "230440005080062", located in the "Centro" district has the values presented in Table 2.

Table 2. Coefficients for the sector "Centro"

β^1	β^2	β^3
370.227	8,834	-2.567

The linear regression equation between each coefficient and income was found, as presented below:

$$\beta^1 = 86{,}102R_i + 0{,}011 \quad (5)$$

$$\beta^2 = 5{,}587R_i + 0{,}0007 \quad (6)$$

$$\beta^3 = -0{,}8217R_i - 0{,}0001 \quad (7)$$

It is noteworthy that the coefficient is the second-order equation *intercept*, which means, the quantity of water demand when the time (t) is zero (the year of 2009). The graphs of Figures 3 to 5 illustrate the aforesaid linear regressions.

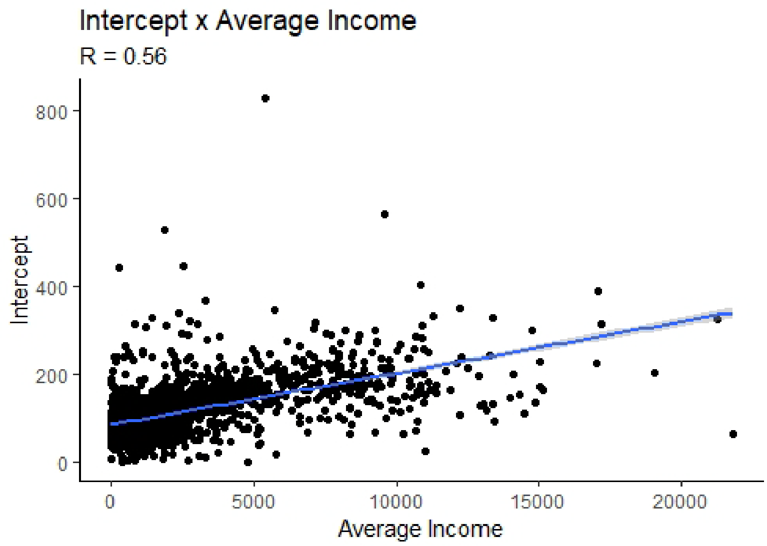

Figure 3. Regression graph of intercepts with average income

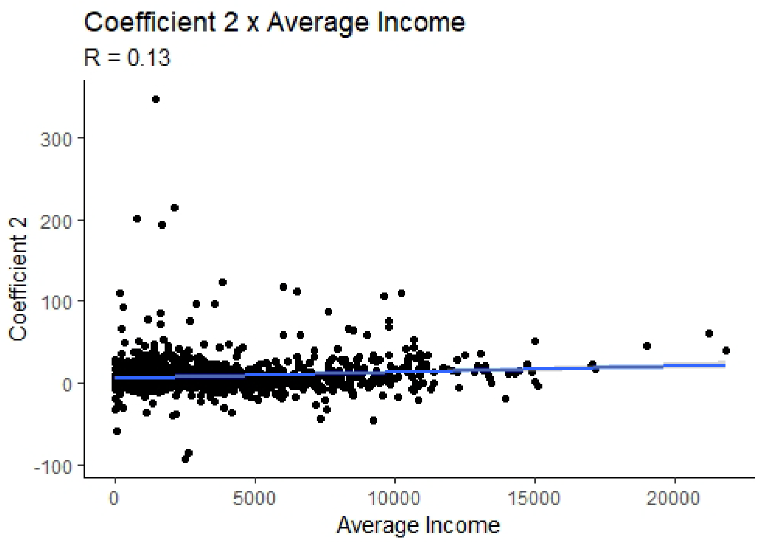

Figure 4. Regression graph of coefficient 2 and average income

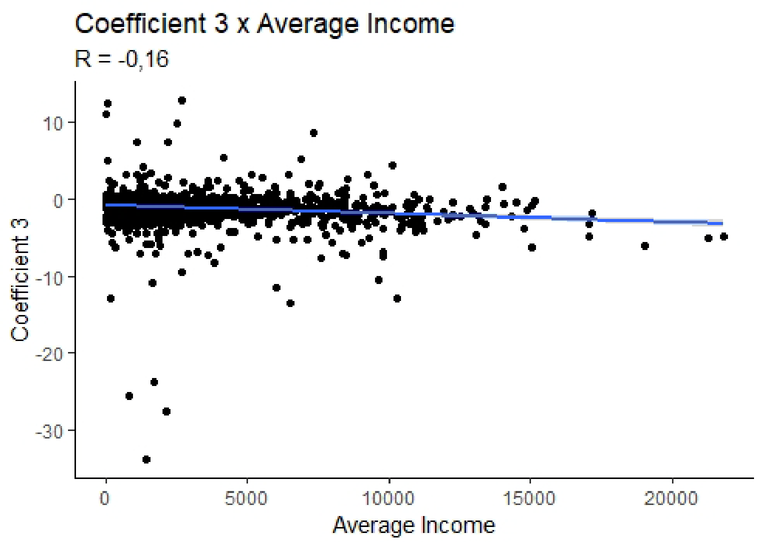

Figure 5. Regression graph between coefficient 3 with average income

Regarding the correlation coefficient (r) for each regression, the relationship presenting a somewhat strong correlation (r = 0.56) is concerning the intercept with average incomes. However, the correlations related to the coefficients 2 and 3 and average incomes are very low (0.13 and -0.16, respectively). Given this situation, it can be inferred that the average income of the sector does not have a solid direct proportional relationship with the intercept of demand, which represents the initial (t = 0) value of water demand. Another point to be evaluated is that the regression lines products of the regressions of coefficients 2 and 3 have a low slope, showing that the coefficients vary in a reduced form by changing the average income. So, it is deduced that the coefficients have similar values and the sectors of Fortaleza in the years 2009 to 2017, overall, presented a similar variation in the water consumption demand.

4. CONCLUSION

The complexity of obtaining the demand-supply balance regarding the water supply in major urban centers becomes more prominent when physical factors are included, such as climate change and drought events, as well as social and economic aspects that guide society's behavior and affect their consumption patterns.

The statistical analysis correlates the urban demand regression coefficients in time with the average monthly income per sector. It was found that the average income does not show a strong dependence on the residential water demand variance. The relationship is stronger with the initial value of the water demand. Regarding the income correlation with the coefficients of time variance (Coefficient 2 and 3), the values obtained were insignificant. Therefore, it is not suitable to estimate the sector water demand using the sector average income because it does not explain the water demand variation coefficients sufficiently.

Due to the higher relation to the initial value of demand, it can be concluded that consumption values per capita increase, according to the growth of HDI and middle-class income. Thus, sectors in the northern part of the city, for example, in neighborhoods like Aldeota and Meireles, present higher consumption per capita.

Finally, a more sophisticated analysis with more extensive number of demand data is recommended, in which the average income considered in each sector is estimated from a time series, not only from a specific year, as was done in this study, which considered only the 2010 average income.

REFERENCES

Cosgrove, W.J., Loucks, D.P., 2015. Water Management: current and future challenges and research directions. Water Resources Research, 51, 4823-4839.

Danielson, L.E., 1979. An analysis of residential demand for water using micro time-series data. Water Resources Research, 15(4), 763-767.

de Maria André, D., Carvalho, J.R., 2014. Spatial determinants of urban residential water demand in Fortaleza, Brazil. Water Resources Management, 28(9), 2401-2414.

Fielding, K.S.. Spinks, A., Russell, S., McCrea, R., Stewart, R., Gardner, J., 2013. An experimental test of strategies to promote voluntary urban water demand management. Journal of Environmental Management, 114, 343-351.

Hanke, S.H., Mare, L.D., 1982. Residential water demand: A pooled, time series, cross section study of Malmö, Sweden. Journal of the American Water Resources Association, 18(4), 621-626.

Li, Q., Schaub, D., 2004. Economic globalization and transnational terrorism: A pooled time-series analysis. Journal of Conflict Resolution, 48(2), 230-258.

Lianqing, X., Yongkun, L., Zhenghang, F., Jieyou, L. 2012. Optimal utilization simulation and decision making system on water resources. Proceeding Environmental Sciences, 12, 1097-1103.

Machado, E.G., 2017. Desigualdades e segregações socioespaciais em Fortaleza, Brasil. O Público e o Privado. 15, 179-207.

Pienaar, G.W., Hughes D.A., 2017. Linking hydrological uncertainty with water resources is equitable allocation decision-making. Water Resources Management, 31, 269-282.

Rothwell, P. M., Eliasziw, M., Gutnikov, et al. 2003. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. The Lancet, 361(9352), 107-116.

Saideman, S. M., Lanoue, D. J., Campenni, M., Stanton, S., 2002 Democratization, political institutions, and ethnic conflict: A pooled time-series analysis, 1985-1998. Comparative Political Studies, 35(1), 103-129.

Schleich, J., Hillenbrand, T., 2009. Determinants of residential water demand in Germany. Ecological Economics, 68(6), 1756-1769.

Stephens, J.M., Sharpe, K., Gore, C., et al. 2018. Core temperature responses to cold-water immersion recovery: a pooled-data analysis. International Journal of Sports Physiology and Performance, 13(7), 917-925.

Suarez-Alminana, S., Peter-Monzonís, M., Walls-Arquiola, J., Andreu, J., Solera, A., 2017. Linking the Pan-European to date is the local scale for global decision making change and water scarcity within water resources planning and management. Science of the Total Environment, 603-604, 126-139.

Ward, M.M., Leigh, J.P., 1993. Pooled time series regression analysis in longitudinal studies. Journal of Clinical Epidemiology, 46(7), 645-659.