Journal of Sustainability Science and Management
Volume 18 Number 1, January 2023: 102-117
eISSN: 2672-7226

MODELLING THE EFFECTS OF SOCIO-ECONOMIC DEMOGRAPHICS ON URBAN WATER USAGE IN KOTA SAMARAHAN, SARAWAK: A NEW EDUCATION HUB IN BORNEO ISLAND

KUOK KING KUOK[1]*, CHIU PO CHAN[2], MD. REZAUR RAHMAN[3], MUHAMMAD KHUSAIRY BAKRI[3] AND CHIN MEI YUN[1]

[1]*Faculty of Engineering, Computing and Science, Swinburne University of Technology Sarawak Campus, Jalan Simpang Tiga, 93350 Kuching, Sarawak, Malaysia.* [2]*Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300 Kota Samarahan, Sarawak, Malaysia.* [3]*Faculty of Engineering, Universiti Malaysia Sarawak, Jalan Datuk Mohammad Musa, 94300 Kota Samarahan, Sarawak, Malaysia.*

**Corresponding author: kkuok@swinburne.edu.my*
Submitted final draft: 19 July 2022 *Accepted: 9 October 2022* http://doi.org/10.46754/jssm.2023.01.007

Abstract: This study was carried out to investigate the influence of socio-economic status on household water usage patterns in Kota Samarahan, which is an education hub in Sarawak, Malaysia. This study commenced with a random sampling of 200 respondents, categorised into low-, medium- and high-income households. The medium-income household category was found to have the highest amount of water usage. The results showed that an increase in income leads to an increase in socio-economic status, dwelling size, and household occupancy. It was also observed that the “numbers of children” influences the increase in water usage within a family. In addition, the data set was further analysed using multiple linear regression modelling (STEPWISE). It was found that an increase in socio-economic demographic factors, including education level, number of female adults, number of clothing washed daily, number of wage earners, and number of dishes washed daily, increased the water usage per household. The findings of this study are crucial to ensuring a sustainable urban water supply in Kota Samarahan.

Keywords: Socio-economic, demographics, Pearson Correlation Coefficient, Multiple Linear Regression Coefficient.

Introduction

Water is essential to all living things on Earth. As for humans, water necessity arises from sheer survival that covers a range of modest daily activities in the household. Third-world countries have faced severe water shortages for decades, especially in urban areas, leading to serious health hazards and fatalities. Economic migration has resulted in the urban population vastly outnumbering the rural population in favour of urbanisation (Alcamo *et al.*, 2007; Avni *et al.*, 2015). Rapid urbanisation had contributed to concentrated housing schemes within the limited urban land, resulting in a considerable upsurge of urban water demands. Additionally, consequent depletion of freshwater resources is apparent across the globe, mainly due to land use and climate changes. The cutting down of plants and trees for urban development has significantly reduced evapotranspiration, infiltration rate, and groundwater runoff (Aly & Wanakule, 2004; Cardell-Oliver, 2014). Malaysia is located within a humid region abundant with rainfall. However, the water stress score is expected to rise from 0.97 in 2020 to 1.78 by 2040 (Maddocks *et al.*, 2015). Therefore, it is crucial to investigate the demographic factors affecting water consumption patterns to ensure a sustainable water supply.

Recognising the relationships between physical factors (e.g., dwelling size, end-use water, presence of landscaping area, occupancy) and sociological factors (e.g., education, socio-economic status) associated with economic factors (e.g., income, water bills) are paramount for determining the influencing factors of water usage (Worthington & Hoffman, 2008; Willis *et al.*, 2013; Ghavidelfar *et al.,* 2017). However, water usage patterns, either through demographic characteristics or their influence, have not been explored and investigated for Kota Samarahan, a new rising education hub of Borneo island, located in Sarawak, Malaysia. This limits the understanding of current water usage patterns in Kota Samarahan areas for adequate and sustainable water supply planning and management. Therefore, this study was conducted to determine the influence of socio-economic demographics and household characteristics on water usage patterns through rigorous statistical analysis and predictive regression modelling by ranking and deducing demographic factors that severely affect water usage patterns. The analysis in this investigation would aid in resource management and crisis mitigation by laying down strategies to counter the steep rise in future demands.

In the past, many studies have been conducted to recognise the influencing factors of water usage. Gomez *et al.* (2019) studied the socio-economic factors affecting water access in rural areas of low- and middle-income countries. The study suggests that gross national income (GNI), female primary completion rate, agricultural activity, rural population growth, and governance indicators, including political stability, corruption control, and regulatory quality, are variables related to water access. Singh and Turkiya (2013) and Hussien *et al.* (2016) also demonstrated positive correlations between the life cycle of the dwelling in relation to the dweller's age, affecting water demand significantly. In addition, Gondo *et al.* (2020) discovered that income is a key determinant of water consumption, followed by affordability, distance to water sources, and the intended use of water in Okavango Delta, Botswana. The model produced by Ashoori *et al.* (2016) highlighted that lower socio-economic status is the most susceptible to price elasticity due to income limits. The developed regression model, which took climate influence into account, demonstrated that higher temperatures and less precipitation had led to a rise in water demand.

However, due to the availability of multiple water sources such as portable water, rainwater, and bottled drinking water, modelling water usage patterns have been substantially less successful, particularly in water-rich developing countries. Such complexities in the infrastructure pose threats to the validation of the accuracy and precision of the collected data. Additionally, the variable and unquantifiable socio-economic, behavioural, cultural, and institutional influences may lead to an incomplete informational circumstance (Yang *et al.,* 2016). Therefore, there is a need to investigate the relationships between physical, social, economic, and demographic factors and water usage patterns in Kota Samarahan, Sarawak, Malaysia. These demographic factors play vital roles in outlining the water demand and usage patterns, which assist the relevant authorities in planning and managing water resources in the coming years.

Study Area

The selected study area is in Kota Samarahan, Sarawak, Malaysia, an education hub for developing countries located on the island of Borneo. It is a suburb located in the southeastern region of the greater metropolitan area of Kuching City, as shown in Figure 1. Kota Samarahan has a population of 128,280 inhabitants, with 16,622 households and 24,512 living quarters spread over 407.1 km^2, accounting for 8.20% of the total land area of the Samarahan Division (Kuok *et al.,* 2011a; Minister's, 2015).

The Samarahan Division's topography is primarily flat and has low-lying areas. In the old days, these areas' main economic activities were coconut, oil palm, and pineapple planting (Kuok *et al.*, 2011b). In the past 25 years, the education-related sectors have become an increasingly important industry in Samarahan. Various higher institutions were set up in the area, including Universiti Malaysia Sarawak (UNIMAS), Universiti Teknologi MARA (UiTM) Kota Samarahan Campus 1 and Campus 2, Tun Abdul Razak Teacher Education Institute, Kota Samarahan Industrial Training Institute (ILPKS) and AAA Zenith Services, which is an English language service provider. Besides, various government training centres have been set up in recent years, including the National

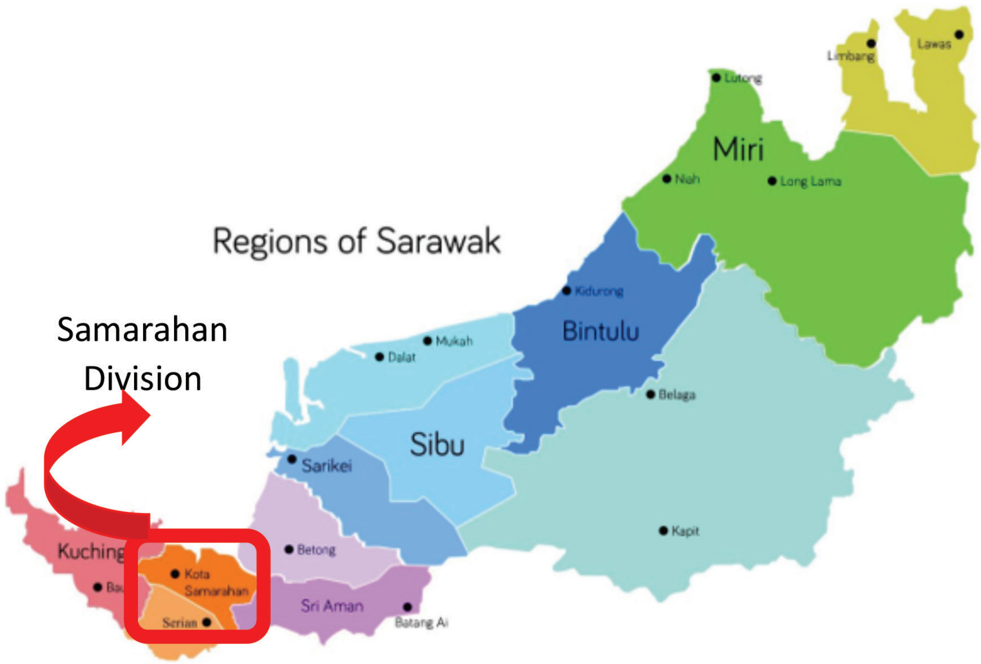

Figure 1: Locality map of Samarahan Division and Kota Samarahan (Louis *et al.,* 2019)

Institute of Public Administration Sarawak Branch (INTAN), Campus Institute of Rural Development Sarawak Branch (INFRA) and Malaysia Air Force Training Centre Sarawak Branch Squadron 330/331 (Kuok & Chiu, 2018).

Meanwhile, a RM400 million Sarawak General Hospital Heart Centre equipped with the latest modern technology and professional cardiac specialists is also located in Kota Samarahan. Commercial centres and residential areas are mushrooming within this emerging educational hub. Therefore, it is important to study the demographic factors affecting the water usage pattern in this area to ensure sustainable water management to cater to the ever-increasing water demand.

Materials and Methods

Data Collection

In March and December 2020, a random sample of Kota Samarahan residents was surveyed using a questionnaire for this study. 50 questions was prepared, and the questionnaire were done in three languages: English, Chinese, and Malay. Participants could respond in the language that is most comfortable to them. The surveys were conducted using three methods: Telephone, email, and face-to-face. The main information collected during the survey include socio-economic demographics, and the sample size was determined using Equation 1.

$$Sample\ Elements = \frac{Z^2 x\ P\ x\ (1-P)}{C^2} \quad (1)$$

where Z is for the confidence level of 99%, P is the standard deviation of 0.5 and C is the confidence interval of 95%. It was calculated that the sample size was 184, and rounded up to 200 respondents for contingencies (Tabachnick & Fidell, 2007). A multiple-choice format dominated the survey questions, which extensively investigated socio-economic status through age, gender, education, income levels, number of wage earners per household, number of occupants per household, number of dependents, and monthly expenditure on the usage of water through average water bills. Physical characteristics of households such as dwelling space, intended tenure, either rent or own, water supply facilities, the presence of an outdoor landscape, the presence of a swimming pool, and a rainwater harvesting tank, were investigated as well. Environmental factors and water usage patterns were also covered by the survey. Data attributing to the frequency and intensity of water end-usage was collected as well. Samplings were carried out at the Summer Shopping Mall, Aiman Shopping Mall, UniMAS, UiTM, Sarawak International Medical Centre (Heart Centre), and households in the vicinity of the universities, including UniGarden, UniVista, and UniCentral.

One questionnaire should take roughly 15 minutes to complete. Only one representative per household is permitted to complete the survey. IBM SPSS Statistics (v. 20) and Microsoft Excel were used to analyse the obtained data. The analysis results will be displayed as pie charts, bar charts, and boxplots.

Development of Statistical Model

The collected data were analysed using IBM SPSS Statistics (v. 20) to assess the strength of the relationship between the independents and dependent variables. The flowchart for the statistical model development is presented in Figure 2.

The developed complex statistical model provided a comprehensive insight into the key demographics affecting water usage in the Kota Samarahan area. Analysis was conducted through the Pearson correlation coefficient (R) and multiple linear regression models (STEPWISE) using Equations 2 and 3, respectively. Implementing STEPWISE linear regressions allows regression modelling of the data, simulating numerous scenarios, whereas SPSS assesses the influence and strength of the independent variables on the dependent variable. The statistical parameters analysed by SPSS include minimum, maximum, distribution of shapes such as kurtosis and skewness, mean, median and mode, and frequencies. These provide the most significant scenario as the inclusion and deletion of variables decrease the inconsistency and variability in terms of predictive accuracy at every iteration.

$$R = \sqrt{\frac{[(r_{y,x1})^2+(r_{y,xn})^2]-(2r_{y,x1},r_{y,xn},r_{x1,xn})}{1-(r_{x1,xn})^2}} \qquad \textbf{(2)}$$

where is the correlation between independent variable "1" and dependent variable is the correlation between independent variable "nth" and dependent variable, and is the correlation between independent variable "1" and independent variable "nth". Thereafter, STEPWISE presented in Equation 3 was employed to predict the value of the dependent variable (Y') based on the values of independent variables (X_1 and X_n).

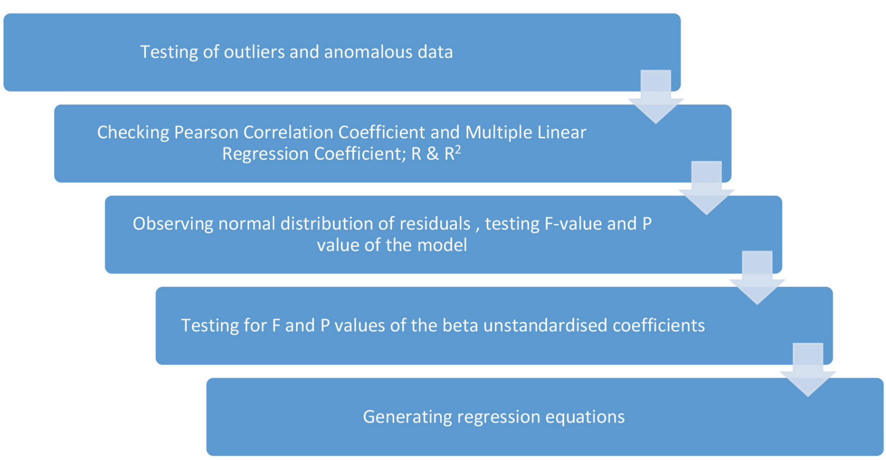

Figure 2: Flowchart for statistical model development

$$Y' = a + b_1X_1 + b_nX_n \quad (3)$$

where a is multiplication regression coefficients with mean values of independent variables and b_1 is calculated using Equation 4.

$$b_1 = \left(\frac{r_{y,x1} - r_{y,xn}r_{x1,xn}}{1-(r_{x1,xn})^2}\right)\left(\frac{SD_y}{SD_{x1}}\right) \quad (4)$$

where SD_y is the standard deviation of dependent variable, SD_{x1} is the standard deviation of independent variable "1", and SD_{xn} is the standard deviation of independent variable "nth". Meanwhile, analysis of variance (ANOVA) tests that relied on the F-test and P values were also carried out to justify whether the grouped variable parameters/impactors are jointly significant for the regression model (Higgins, 2006; Uyanık & Güler, 2013). The imperative for testing whether the null hypothesis exists within a data set was carried out using Equation 5.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k \quad (5)$$

where μ represents the mean of the group, k represents the number of groups tested under the ANOVA test. The analysis of the extracted data was based on the following statistical parameters:

- Null hypothesis, H_o. There is no correlation between the impactors and water usage. Therefore, for regression models H_o: $R^2 = 0$.
- Alternate hypothesis, H_1. Correlation between the impactors and water usage exists. Therefore, for regression models H_1:$R^2 \neq 0$.
- Statistical significance, P-value, $\rho \leq \alpha$. Therefore, $\rho_{significance} \leq 0.05$, at alpha level of the model tests the significance at 5%, proving 95% of the model is statistically significant and inconsistent with the null hypothesis, H_o.
- Variation acceptability, F-value, $f \geq 1.0$ to maintain inconsistency with the equal ratio of variation between sample means and variation within the samples.
- Coefficient of Variability, $C_{vreg} \leq 10\%$.
- Pearson Correlation coefficients; significant at H_1:$R^2 \neq 0$ and insignificant at H_o: $R^2 = 0$.

The regressions were modelled on the influence of all the impactors surveyed, including physical, social and economic demographic factors. "Water Bills" is the dependent variable due to its approximate representation of average monthly water usage within a household, as higher water bills are a result of higher water usage, and vice versa.

Results and Discussion

The analysis of the household characteristics revealed that out of the 200 respondents, 64.31% lived in landed houses, while 36.87% lived in apartments and condominiums, as shown in Figure 3. Figure 4 shows the composition of family members, and the average occupancy rate is 3.84 per household. The average number of rooms per household ranges from 3 to 5, as presented in Figure 5. The survey found that the average number of years respondents reported living within the survey households is 8.209.

Regarding education levels, 45.49% of the respondents have graduated from high school and 45.49% are holding a first degree. Out of the 200 sampled responses, only 9.02% were recorded to hold postgraduate degrees, as presented in Figure 6. The income breakdown in Figure 7 illustrated that most respondents are from the low-income group, earning RM500 to RM5,000 per month, covering 59.6% of the sample population. As seen in Figure 8, 43.14% of the respondents have two wage earners in a family, followed by 33.33% being single-wage earner families, and 19.22% for three of more wage earners in a family.

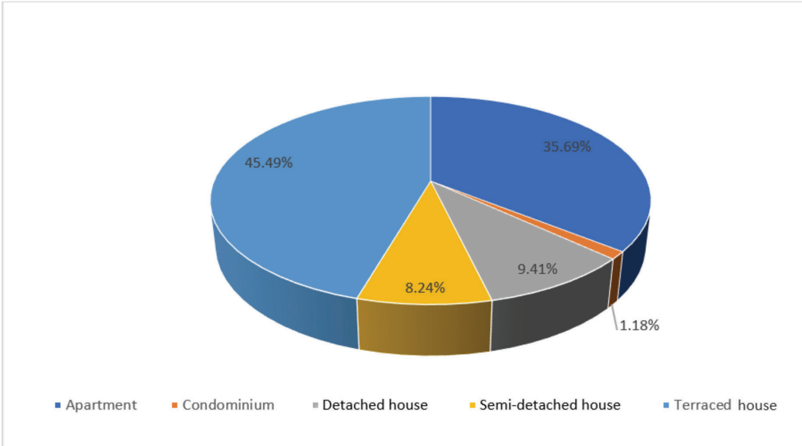

Figure 3: Types of residence

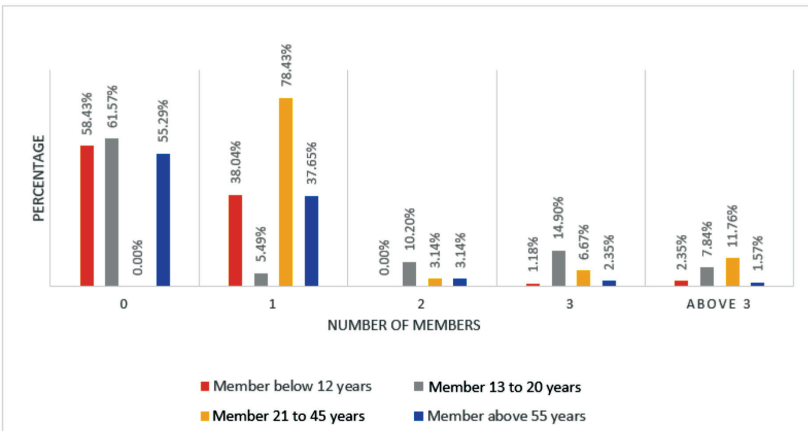

Figure 4: Family composition

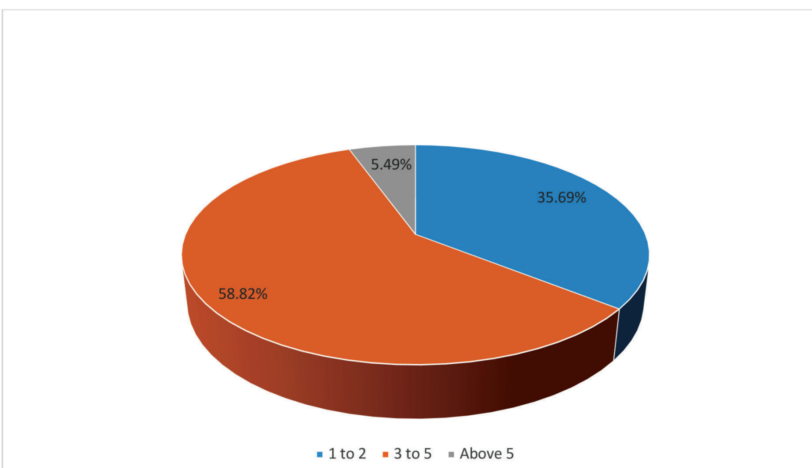

Figure 5: Number of rooms

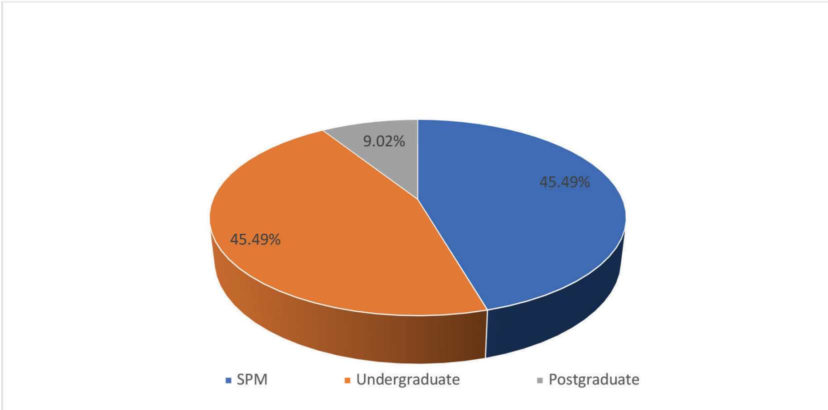

Figure 6: Education level breakdown

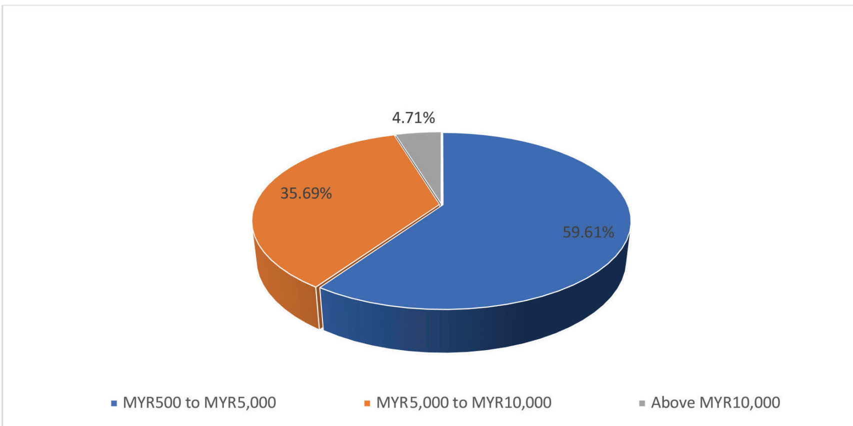

Figure 7: Income breakdown

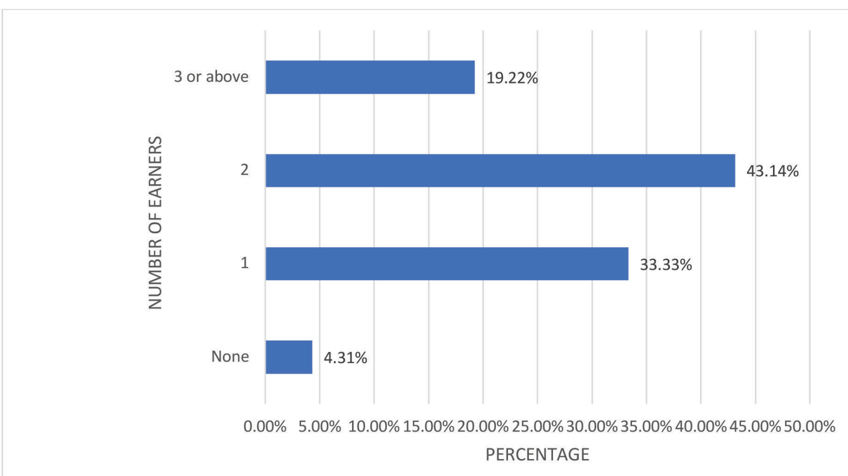

Figure 8: Number of earners per household

Influence of Socio-economic Status on Water Bills

The basic defining variable of a household's socio-economic status is rooted in household income. For the analysis, the responses were segmented into three income groups, namely low, medium and high, as summarised in Table 1. It is evident that as the income segments spread from low to high, the average water bill increases, and so does the average household occupancy by an insignificant margin.

Thereafter, the strength of the relationship between the economic status and the water bill was further assessed using the Pearson correlation coefficient (R). It was found that household income increases with the increase in reported education level within the household, and this relationship possesses a correlation R of 0.616. The boxplot in Figure 9 illustrates the minimum, lower quantile, median, upper quantile and maximum income for SPM, undergraduate and postgraduate education levels.

The boxplot whiskers show that respondents with a SPM-level education mostly earn RM500 to RM5,000, categorised under the low-income segment. The income for respondents with undergraduate education levels spreads across all income segments. Still, it remains skewed, mainly between the low- and medium-income segments, and the median income falls within the range of RM5,000 to RM10,000. The median income for postgraduate respondents also falls within the range of RM5,000 to RM10,000 but is mostly skewed towards the medium- and high-income segments. Based on the correlation coefficient and the boxplot analysis, it is concluded that education level influences income as higher education levels lead to higher income.

Higher education results in an increase in income, which subsequently may lead to an increase in the dwelling size, such as from an apartment to a landed house. This relationship possesses an R coefficient of 0.423, which signifies a positive linear relationship between income level and household type. Figure 10 indicates that the median income for those who own an apartment is within the range of RM500 to RM5,000 and the median income from owning a landed house is within the range of RM5,000 to RM10,000.

It should be noted that the influence of education on income and household type indicates that the increase in socio-economic status leads to a more prosperous and extravagant lifestyle. It is evident in Figure 11 that as the income level increases, water bills increase as well. This positive linear relationship is governed by an R coefficient of 0.406. The results also revealed that the median monthly water bills for the low-income group are RM18, with a lower quantile of RM15, and upper quantile of RM30. The median, lower and upper quantiles of monthly water bills for the middle-income group are RM30, RM26 and RM38, respectively. Meanwhile, the monthly water bills for the high-income group are RM47, RM32 and RM55 for the median, lower and upper quantiles, respectively.

The R coefficient is also used to investigate the relationship between the number of rooms and water bills. This relationship consistently yields a positive relationship with R=0.475, showing that an increase in the number of rooms

Table 1: Income segments

Segment	Income	Frequencies of Responses	Average Water Bills	Average Household Occupancy
Low	RM500 to RM5,000	152.00	RM23.62	3.45
Medium	RM5,000 to RM10,000	91.00	RM32.25	3.82
High	Above RM10,000	12.00	RM41.42	4.25

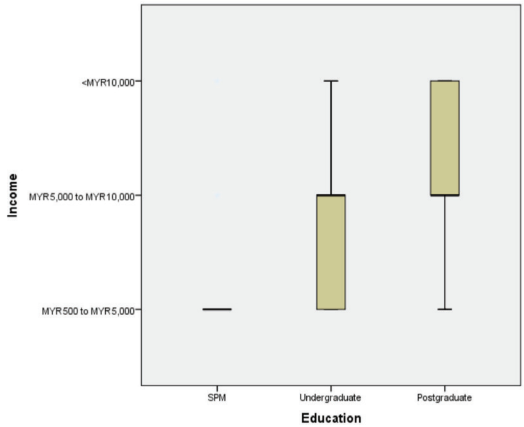

Figure 9: Boxplot of income and education

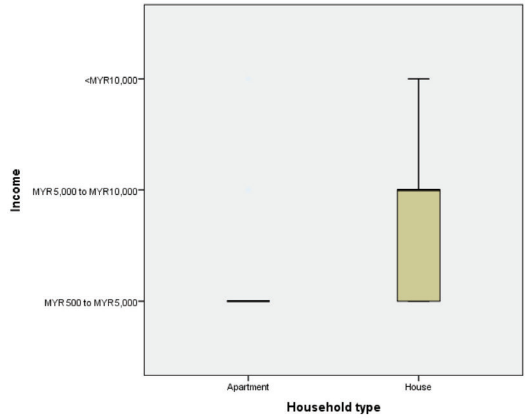

Figure 10: Boxplot of income and household type

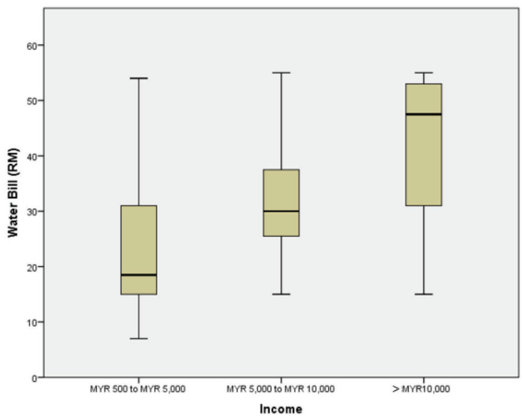

Figure 11: Boxplot of water bills and income

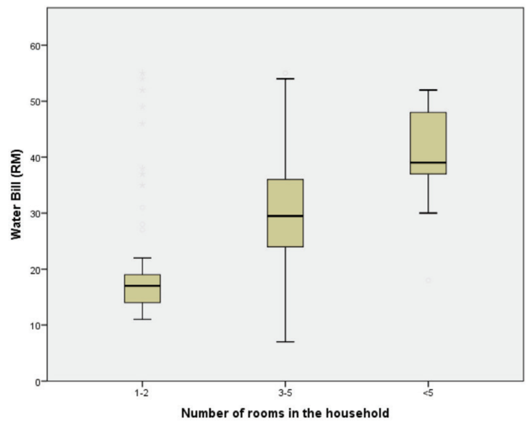

Figure 12: Boxplot of number of rooms and water bills

in a household led to an increase in water bills, as presented in Figure 12. The median, lower and upper quantiles for the monthly water bills for a household with 1 to 2 rooms is found to be RM17, RM14 and RM19, respectively. The monthly water bills for a household with 3 to 5 rooms are RM30 for the median, RM24 for the lower quantile, and RM36 for the upper quantile. For households with more than 5 rooms, the monthly median water bill was RM39, the lower quantile was RM37 and the upper quantile was RM48.

The relationship between the number of family members and water bills was also investigated. The R coefficient of 0.375 shows significant evidence of a positive relationship between occupancy and water bills. Figure 13 presents the median, lower and upper quantiles of monthly water bills for a household with 1 to 2 occupants are RM17, RM15 and RM19, respectively. As the number of occupants increased to 3 to 4 persons, the monthly water bills increased in parallel, with median=RM28, lower quantile=RM24 and upper quantile=RM37. The monthly water bills for a household with 4 to 5 occupants were RM35 for the median, RM25 for the lower quantile and RM38 for the upper quantile. The median monthly water bill for a household with more than 5 members is RM35. The upper quantile of monthly water bills may be RM50 for a family house, whereby the lower quantile is only RM17 for a household with more than 5 occupants. The lower quantile is relatively low, probably due to it being a rented house and occupants do not stay in the house throughout the day.

The R coefficient can be further assessed by how water bills correlate to the breakdown of family composition, including the number of female adults, male adults, children, and the elderly. With a R coefficient of 0.368, it can be inferred through observation that the number of female adults in the household establishes a positive linear relationship with water bills. The boxplot analysis in Figure 14 shows that the median, lower and upper quantiles of monthly water bills for 1-2 female adults are RM25, RM17 and RM34, respectively. As the number of female adults increased to 3-4, the median, lower and upper quantiles of monthly water bills increased to RM26, RM15 and RM45, respectively. The median monthly water bills increased to RM35 as the female adults increased to 4-5, and RN37 with more than 6 female adults.

Meanwhile, a positive linear relationship has also been established between the number of male adults and water bills, with a R coefficient of 0.326, which is similar to the relationship between the number of female adults and water bills. Figure 15 illustrates an increase in water bills in the inter-quantile range and median as the number of male adults rises, and the values obtained are close to the influence of water usage by female adults.

Figures 16 and 17 present the influence the number of children and elderly have on water bills, respectively. The boxplot demonstrated that the median monthly water bill is the same

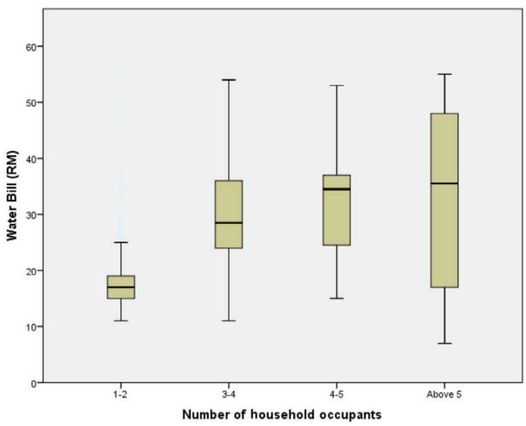

Figure 13: Boxplot of water bills and occupancy

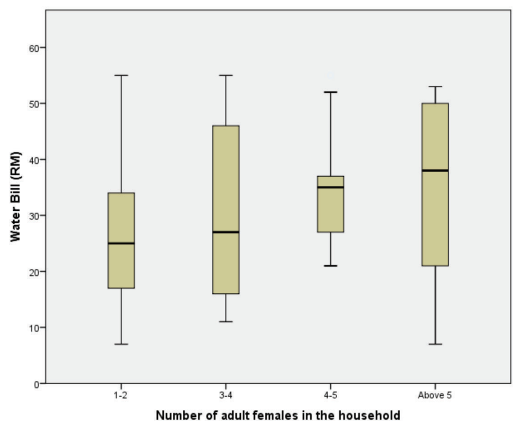

Figure 14: Boxplot of water bills and number of female adults

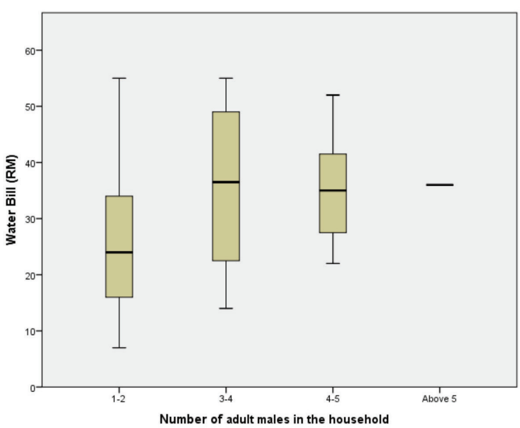

Figure 15: Boxplot of water bills and number of male adults

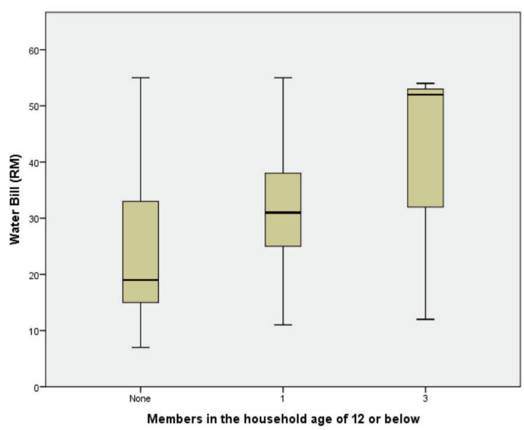

Figure 16: Boxplot of water bills and number of children

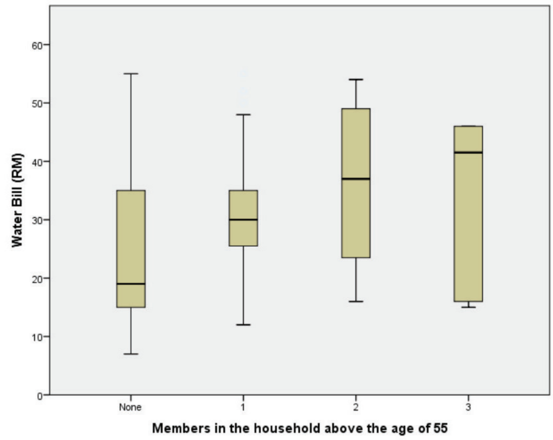

Figure 17: Boxplot of water bills and number of elderlies

for one child under 12 and one elderly in a family, which amounted to RM31. As children under 12 increased to 3, the median monthly water bills increased significantly to RM54. In contrast, the median monthly water bills for 3 elderlies aged above 55 only amounted to RM43. These statistical results proved that children's influence on water usage is more significant than the elderly.

Table 2 illustrates the breakdown of the micro-components survey of residential indoor water consumption in Kota Samarahan. The low-income group was reported to have the lowest number of all micro-components within the sample population. In comparison, the medium-income group was recorded to have the highest number of micro-components within the household. This is consistent with the findings of numerous studies that the medium-income group tends to overspend and live extravagantly. The high-income group was reported to have marginally higher micro-components than the low-income group, but much lower than the medium-income group.

The results show that the average micro-components per household were 3.70 water taps, 0.57 bathtubs, 2.32 hand wash basins, 2.31 flushing toilets, 2.30 showers, 0.64 dishwashing machines, and 1.12 top-loading washing machines. The frequency usage data obtained from the respondents provide general insight into how often the micro-components were utilised. The collected data were further developed into a statistical model for analysing the influence of the micro-components' water usage on water bills.

Table 2: Micro-components of water end usages according to income group

	Low-income Group	**Medium-income Group**	**High-income Group**	**Average**
Number of water taps per household	3.34	4.33	3.42	3.70
Number of bathtubs per household	0.30	0.75	0.67	0.57
Number of hand wash basins per household	1.93	3.11	1.92	2.32
Number of flushing toilets per household	1.62	3.04	2.25	2.31
Number of showers per household	1.72	3.00	2.17	2.30
Number of dishwashing machines per household	0.39	0.78	0.75	0.64
Number of top-loading washing machines per household	1.14	0.96	1.25	1.12

Table 3 presents the breakdown of frequency usage of the micro-components. The bathing choice for all three income groups is mainly recorded as showers, instead of bathtubs. The highest bathing frequency was reported for the low-income group, most probably due to their highly intense jobs that required physical involvement. The second highest bathing frequency is from the medium-income group and lastly, the high-income group. The survey results also indicated that the bathing time is reduced from "20 to 30 minutes" for the high-income group to "10 minutes or lesser" for the low-income group.

The frequency of using laundry for the low-income group ranges between twice to thrice a week, while the medium-income group was closer to once a week. This is explainable as low-income respondents are assumed to have physically intense jobs, requiring regular or higher frequency of clothes laundering. On the other hand, the high-income group tends to have clothes laundering daily or twice a week. Although the lower-income group tends to show higher frequencies of clothes laundering, the number of cycles per laundry typically is only 1 to 2. The medium-income group tends to lean towards 2 to 3 cycles per laundry. The high-income group requires 1 to 2 cycles per laundry with moderate frequencies.

For dishwashing, the low-income group prefers to rely on washing the dishes by hand. The medium- and high-income groups skew towards the use of dishwashing machines mixed with hand washing. The number of dishes for washing is between 6 to 15 per day. The results show that the average number of dishes washed was between 6 to 10 per day for the low-income group, while the medium- and high-income groups were towards the upper range of 15 dishes per day.

Outdoor irrigation has been a crucial aspect of water usage across the globe. Although Malaysia experiences sufficient rainfall, it is observed that outdoor landscapes are irrigated manually through water hoses and automatic sprinklers for the high-income group. The medium-income group, meanwhile, remains consistent with costly approaches, such as automatically timed sprinklers without soil moisture, and rain sensors, which are cost-effective and conserve water in the long run. Meanwhile, the low-income group leans toward ignoring the outdoor landscape if present.

The frequency of washing sidewalks, driveways, garages, and entire households for the low-income group remains between once to twice a week. the medium-income group usually washes sidewalks and the entire household once a week, but washes the driveways and garages up to twice or thrice a week. Responses from the high-income group are much more consistent with cleaning the entire household at least twice a week. Both high- and low-income groups were

Table 3: Frequencies of water end-uses according to income group

	Low-income Group	**Medium-income Group**	**High-income Group**
Frequency of washing household per week	1.02	1.20	2.17
Frequency of washing garage per week	1.07	2.35	2.58
Frequency of washing driveway per week	1.06	2.36	2.58
Frequency of washing sidewalks per week	1.07	1.07	2.75
Frequency of outdoor landscape irrigation per week	0.10	2.21	2.42
Loads of dishwashing per day	8.11	15.00	15.00
Frequency of laundry per week	2.51	1.10	1.50
Frequency of bathing per day	2.25	2.03	1.83

reported to wash their vehicles at home at a frequency of twice a month, slightly lower than the medium-income group, with a frequency of between "twice a month" and "once a week".

The comparison of water end-uses among all the income groups provides a consistent and clear view of how the micro-components for the medium-income group remain water-intensive. These findings are consistent with various studies. Apart from household income, other factors affecting water usage are household types and locality. Smaller households tend to have higher water demands for indoor usage, whereas larger households require water for outdoor maintenance.

The household characteristics and average monthly bills collected from 200 households were further utilised for statistical analysis using the STEPWISE multiple linear regression method. The STEPWISE linear regression was used to measure the strength of the influence of independent input variables on the dependent variable "water bills" under different scenarios using IBM SPSS (v. 20). The analysis of the regressions requires the models to fit into the parameters for rejection of the null hypothesis. The input variables in the regression model are as follows:

- Socio-economic demographics: Earners, education, income
- Household characteristics: Household type, number of years lived in the household, number of rooms in the household, presence of outdoor landscape
- Family composition: Members in the household aged 12 years or below, members in the household between the age of 21 years and 45 years, members in the household above the age of 55 years, number of adult females in the household, number of adult males in the household
- Water end-usages for each household: The number of top-loading washing machines, the number of dishwashing machines, the number of showers, the number of flushing toilets, the number of hand wash basins, number of bathtubs, number of water taps, frequency of laundry in a week, number of cycles of laundry in a single wash, the number of dishes washed per day
- Personal hygiene: Bathing frequency and duration per day
- External uses: Place and frequency of vehicle wash, source of outdoor landscape irrigation, frequency of washing sidewalks, frequency of washing driveway, frequency of washing household

The histogram in Figure 18 illustrates the distribution of the regression standardised residuals against the frequency of occurrence. The "bell-shaped" curve in the histogram indicates the normal distribution of the errors. Meanwhile, the normal P-P plot in Figure 19 illustrates the distribution of the observed cumulative probability against the expected cumulative probability. The objective of the outputs is to verify whether the observed residual line or errors within the collected data displays significant variations with the errors incurred through predictive values obtained through the regression equation. In this study, it was found that the variation observed is negligible. Therefore, even though the bell curve is slightly skewed towards the left, the model still fits the normal distribution.

Six models were generated by IBM SPSS (v. 20) through STEPWISE linear regressions for the data analysis, including the low-, medium- and high-income groups. The correlation coefficient for the regression model R^2=0.550 and R=0.741 proved that the dependent variable "water bills" can be predicted up to 55% by the independent variables identified by IBM SPSS (v. 20). As the variables are measured with their respective weights, the water bill was predicted using Equation 6.

$$(MYR)\ Y = 2.28 + 11.959\,X_1 + 3.559\,X_2 + 4.78\,X_3 + 3.147\,X_4 + 2.445\,X_5 + 3.375\,X_6 \quad \textbf{(6)}$$

where Y is the water bill (RM), X_1 is income, X_2 is the number of female adults in the household, X_3 is the laundry medium, X_4 is the number

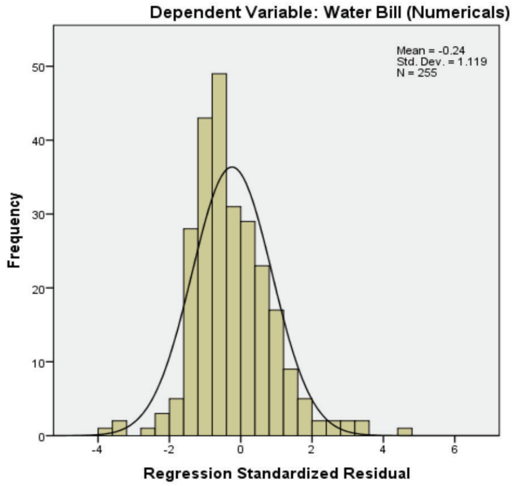

Figure 18: Normal probability distribution plot

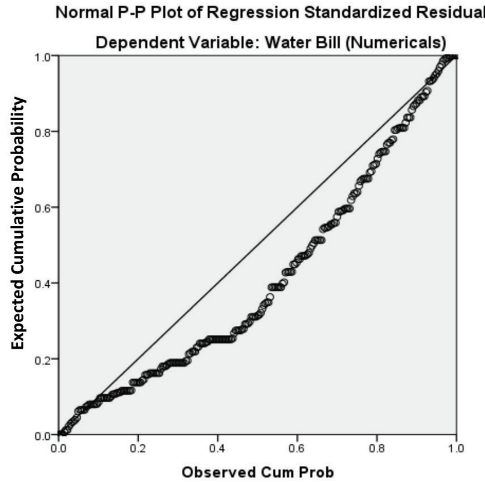

Figure 19: Normal P-P plot for regression residuals

of earners, X_5 is the number of dishes washed per day, and X_6 is the education level. Here the weighted unstandardised beta coefficients represent the gradients of each independent variable's relation to the dependent variable. For the regression prediction, it was observed that:

- An increase of income level within the household from low to medium, and from medium to high, would result in an increase of RM11.959 in the monthly water bills.
- An increase of one female adult within the household would result in an increase of RM3.559 for the monthly water bills.
- Washing the clothes inside the household either using the washing machine or hand washing would result in an increase of RM4.784 in the monthly water bills.
- A unit increase in the number of earners within the household would result in an increase of monthly water bills by RM3.147.
- An increase of one number of dishes washed per day would result in an increase of RM2.445 in the monthly water bills.
- An increase in the education level, from SPM high school to undergraduate, from undergraduate to postgraduate would increase the monthly water bills by RM3.375.

The predictions discussed are based on the ideal scenario, where the increase was influenced only by the variables under consideration while others remain constant. The strongest impactor was household income for the entire data set. Furthermore, the influencing input independent variable "education" has a linear relationship with the dependent variable "water bills". Higher incomes encourage people to live more opulently and prosperously, which may lead to an increase in home size, from an apartment to a landed house, or from a terrace house to a semi-detached house. This, in turn, requires more water for activities such as washing and maintaining the lawn within the housing compound.

Conclusion

Through vigorous statistical analysis and development of the Pearson correlation coefficient(s) across all the significant impactors, it was observed that higher education levels will lead to higher household incomes. The increase in household income allows people to opt for a larger dwelling size. As the dwelling size increase, the correlation coefficients illustrated a positive relationship with the increase in the number of rooms, which further correlated positively to the incease in household occupation. This relationship was then extended to water usage, which consistently increased water bills with the increase in income, dwelling size, number of rooms, and household occupancy. Water end-usage and micro-component utilisation was found to be consistent with the findings of the literature, where the medium-income group tends to be resource intensive.

The findings of correlations were put through the STEPWISE multiple linear regression models. It was found that the main impactors that contribute to the dependent variable "water bills" are income, the number of female adults in the household, the laundry medium, the number of wage earners, the number of dishes washed per day and education level.

In conclusion, statistical significance parameters for modelling the data were identified. Therefore, the study results are conclusive of the hypothesis that socio-economic demographics have an influence on physical characteristics, eventually increasing water usage within a household. Higher income constitutes high socio-economic status, which leads to a more extravagant lifestyle, leading to an increase in water usage. Hopefully, all these data will be used to determine the average water demand and future water supply planning and management, particularly for this rising education hub in Sarawak, Malaysia.

Acknowledgements

The authors would like to extend our gratitude to all the interviewees that had helped and participated in this research study.

References

Alcamo, J., Flörke, M., & Mäerker, M. (2007). Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences Journal*, *52*(2), 247-275. https://doi.org/10.1623/hysj.52.2.247

Aly, A. H., & Wanakule, N. (2004). Short-Term forecasting for urban water consumption. *Journal of Water Resources Planning and Management*, *130*(5), 405-410. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:5(405)

Ashoori, N., Dzombak, D. A., & Small, M. J. (2016). Modeling the effects of conservation, demographics, price, and climate on urban water demand in Los Angeles, California. *Water Resources Management*, *30*(14), 5247-5262. https://doi.org/10.1007/s11269-016-1483-7

Avni, N., Fishbain, B., & Shamir, U. (2015). Water consumption patterns as a basis for water demand modeling. *Water Resources Research*, *51*(10), 8165-8181. https://doi.org/10.1002/2014WR016662

Cardell-Oliver, R. (2013). Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resources Research*, *49*(12), 8589-8599. https://doi.org/10.1002/2013WR014458

Chief Minister Department. (2015). *Sarawak Facts and Figures: Sarawak State Planning Unit*. http://www.spu.sarawak.gov.my/

Ghavidelfar, S., Shamseldin, A. Y., & Melville, B. W. (2017). A multi-scale analysis of single-unit housing water demand through integration of water consumption, land use and demographic data. *Water Resources Management*, *31*(7), 2173-2186. https://doi.org/10.1007/s11269-017-1635-4

Gomez, M., Perdiguero, J., & Sanz, A. (2019). Socioeconomic factors affecting water access in rural areas of low and middle income countries. *Water*, *11*(2), 202. https://doi.org/10.3390/w11020202

Gondo, R., Kolawole, O. D., Mbaiwa, J. E., & Motsholapheko, M. R. (2020). Demographic and socio-economic factors influencing water governance in the Okavango Delta, Botswana. *Scientific African*, *10*, e00602. https://doi.org/10.1016/j.sciaf.2020.e00602

Higgins, J. (2006). *The radical statistician: Unleashing the power of applied statistics in the real world*. Higgins Group Publishing.

Hussien, W. E. A., Memon, F. A., & Savic., D. A. (2016). Assessing and modelling the influence of household characteristics on per capita water consumption. *Water Resources Management*, *30*(9), 2931-2955. https://doi.org/10.1007/s11269-016-1314-x

Kuok, K. K., Harun, S., & Chan, C. P. (2011b). Investigation best number of tanks for hydrological tank model for rural catchment in humid region. *Journal the Institution of Engineers, Malaysia*, *72*(4), 1-11. http://dspace.unimap.edu.my/xmlui/bitstream/handle/123456789/18973/2.investigation.pdf?sequence=1

Kuok, K. K., Harun, S., & Chiu, P. C. (2011a). Comparison of particle swarm optimization and shuffle complex evolution for auto-calibration of hourly tank model's parameters. *International Journal of Advances in Soft Computing and its Applications*, *3*(3), 1-17.

Kuok, K. K., & Chiu, P. C. (2018). Indigenous drinking-water consumption pattern of residents in Kuching city: Results of a pilot study. *Journal of Water, Sanitation and Hygiene for Development*, *8*(4), 817-824. https://doi.org/10.2166/washdev.2018.004

Louis, Y. H. T., Kuok, K. K., Imteaz, M., Lai, W. Y., & Derrick, K. X. L. (2019). Development of whale optimization neural network for daily water level forecasting. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(3), 354-362. https://doi.org/10.30534/ijatcse/2019/04832019.

Maddocks, A., Young, R. S., & Reig, P. (2015). Ranking the world's most water-stressed countries in 2040. *World Resources Institute*. https://www.wri.org/blog/2015/08/ranking-world-s-most-water-stressed-countries-2040

Singh, O., & Turkiya, S. (2013). A survey of household domestic water consumption patterns in rural semi-arid village, India. *GeoJournal*, *78*(5), 777-790. https://doi.org/10.1007/s10708-012-9465-7

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, *106*(1), 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

Willis, R. M., Stewart, R.A., Giurco, D.P., Talebpour, M.R., & Mousavinejad, A. (2013). End use water consumption in households: Impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production*, *60*(1), 107-115. https://doi.org/10.1016/j.jclepro.2011.08.006

Worthington, A. C., & Hoffman, M. (2008). An empirical survey of residential water demand modelling. *Journal of Economic Surveys*, *22*(5), 842-871. https://doi.org/10.1111/j.1467-6419.2008.00551.x

Yang, T., Shi, P., Yu, Z., Li, Z., Wang, X., & Zhou, X. (2016). Probabilistic modeling and uncertainty estimation of urban water consumption under an incompletely informational circumstance. *Stochastic Environmental Research and Risk Assessment*, *30*(2), 725-736. https://doi.org/10.1007/s00477-015-1081-x