International Journal of Statistics and Applied Mathematics 2019; 4(6): 128-132

International Journal of Statistics and Applied Mathematics

**ISSN:** 2456-1452
Maths 2019; 4(6): 128-132

www.mathsjournal.com
Received: 10-09-2019
Accepted: 14-10-2019

**Robert Orwenyo Onyoni**
Department of Statistics and Actuarial Science,, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

**George Otieno Orwa**
Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

**Zablon Maua Muga**
Department of Applied Statistics, Financial Mathematics and Actuarial Science Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

**Corresponding Author:**
**Robert Orwenyo Onyoni**
Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

# Modeling household water demands using sinusoidal models

**Robert Orwenyo Onyoni, George Otieno Orwa and Zablon Maua Muga**

**Abstract**
Household water in Kenya is used in agricultural activities, industrial activities and other uses. A lot of water is consumed by indoor appliances. The water management strategy affects the household demand for water. The future demand of water in Kenyan towns has remained uncertain. Therefore this study sought to model household water demand using the non- parametric model in the spectral domain. This study largely relied on the secondary data that was collected from Gusii Water and Sanitation Company (GWASCO) situated in Kisii town. This study inspected the data for trend by applying Mann Kendal trend test and it was realized that the p-value = 0.012< Siginificance level, α = 5% indicating that the water demand data had a trend. The parameters of the sinusoidal model was estimated and they were: $B_1 = 90.28176, B_2 = 69.06094 \ and \ \mu = 3586.94540$ The developed model was used in conducting in sample forecasting. The forecasted values were plotted together with those of GWASCO and comparison was made. This study noted no significant difference in the plots. Therefore, the model was employed in conducting out of sample forecasting. The out sample forecasted values increased with time. This study concluded that the water data had a trend, had periodic variations hence could be modeled using the sinusoids and that the forecasts increased with time.

**Keywords:** Trend, spectral domain, non-parametric, sinusoidal model

## 1. Introduction

Water is considered to be one of the crucial resources in the survival of living things. Human beings rely on water to conduct household chores. The location, function and household personal preferences are the main contributors in the determination of household water demand. The quantity of water demanded per capita has been provided by various organizations. According to Kenya Ministry of Water Resource Management and Development, the country's water supply stands at 647$M^3$ per capita. World Resource Institute estimates the volume of water demanded to be 590$M^3$ per capita. This indicates a huge variation in the estimates by various institutions [1, 2].

The procedure of unraveling the quantity of water demanded in a country relies heavily on the population estimates. The Kenyan population estimates have been on the upward trend indicating an ever increasing demand of water in the country. This has led to the development of institutions in various parts of the country to help in the management of the household water demand [2].

In statistics, frequency domain refers to the analysis of the mathematical functions/signals with respect to the frequency rather than time [3]. The detail of determining the spectral density of a sample of population data is very old; this was applied by the first astronomers who tried to study the length of the year and the period of the moon. This entailed the determination of the maxima and/or estimation of the period [4].

The first significant frequency estimation problem took place in the 19$^{th}$ century where the probability theory and the use of Fourier transform came into being. In the frequency estimation process, the model might contain a single cosine with amplitude, phase and a frequency contaminated by the noise error with unknown variance. Basically a researcher is interested in the amplitude, phase or noise variance [4, 5].

## 2. Material and Methods

### 2.1 Cox-Stuart Trend Tests

According to [5], C-S test is normally employed to detect the non-random pattern, that is the periodic patterns. C-S test compares the first half and the second half of the sample data. When the data has a down ward trend, the observatons in the first half are expected to be greater than the observation in the second half. When the data has upward trend, the observations in the first half are expected to be smaller than the observations in the second half. When the data in question has no trend, then the researcher should expect smaller differences between the two halves of the sample data due to the randomness of the data. To perform the C-S test, the sample differences are computed as shown below;

$$Y_1 = x_{1+m} - x_1, Y_2 = x_{2+m} - x_2, Y_3 = x_{3+m} - x_3, \ldots, Y_m = x_n - x_{n-m} \quad (3.1.1)$$

Where $m = \frac{n}{2}$ if and only if n is even and $m = \frac{(n+1)}{2}$ if and only if n is odd.

The differences which are equal to zero are ignored in this case. Denoting the sample data with positive differences by $Y_1$, $Y_2$, $Y_3$, …, $Y_m$; then the C-S test is a sign test applied to the sample data of non-zero differences $Y_1$, $Y_2$, $Y_3$, …, $Y_m$. Let $sgn(b) = 1$, if $b > 0$ and $sgn(b) = -1$, if $a < 0$. Then the C-S statistic is given as:

$$C - S\ Statisic = \sum_{j=1}^{m} sgn(Y_j) \quad (3.1.2)$$

The hypotheses tested at the significance level (α) are:
$H_0$: There is no periodic trend in the household water data
$H_A$: There is periodic trend in the household water data

### 2.2 Sinusoidal modeling

According to [6], the periodicities were modeled using the sinusoidal model of the form:

$$Y_t = \mu + A * \sin\left(\frac{2\pi*12}{96} * t + P\right) + \varepsilon_t \quad (3.2.1)$$

In this equation $\mu, A$ and $P$ cannot be estimated by the standard least square method. To achieve this estimator we expanded as shown below;

$$A\sin\left(\frac{\pi}{4} * t + P\right) = A\sin\left(\frac{\pi}{4} * t\right)\cos(P) + A\cos\left(\frac{\pi}{4} * t\right)\sin(P) = B_1\sin\left(\frac{\pi}{4} * t\right) + B_2\cos\left(\frac{\pi}{4} * t\right) \quad (3.2.2)$$

Where $B_1 = A * \cos(P)$ and $B_2 = A * \sin(P)$
Hence the structure of the sinusoidal model becomes:

$$Y_t = \mu + B_1\sin\left(\frac{\pi}{4} * t\right) + B_2\cos\left(\frac{\pi}{4} * t\right) + \varepsilon_t \quad (3.2.3)$$

Where the parameters $\mu, B_1$ and $B_2$ can be estimated by the method of least squares estimation process.
It can be seen that,

$B_1 = A\cos(P)$, therefore, $B_1^2 = A^2 cos^2(P)$ and

$B_2 = A\sin(P)$, therefore, $B_2^2 = A^2 sin^2(P)$

$$B_1^2 + B_1^2 = A^2\cos^2(P) + A^2\sin^2(P) = A^2[cos^2(P) + \sin^2(P)] \quad (3.2.4)$$

But $cos^2(P) + \sin^2(P) = 1$. This implies that $B_1^2 + B_1^2 = A^2$

Therefore the Amplitude (A) can be estimated by

$$A = \sqrt{B_1^2 + B_2^2} \quad (3.2.5)$$

Also $\frac{B_2}{B_1} = \frac{\sin(P)}{\cos(P)} = \tan(P)$, Hence the $P = \tan^{-1}\left(\frac{B_2}{B_1}\right)$

In some cases the amplitudes of the peaks and the troughs are rarely equal, then the sine wave model can be re-written using the Fourier series model:

$$Y_t = \mu + \sum_{i=1}^{h} A_j \sin(2\pi * f_i * t + P_i) + \varepsilon_t$$

$$= \mu + \sum_{i=1}^{h} B_{1i}\sin(2\pi * f_i * t) + \mu + \sum_{i=1}^{h} B_{2i}\cos(2\pi * f_i * t) + \varepsilon_t \quad (3.2.6)$$

Equation (3.2.6) above entails a linear combination of the harmonic waves (HW). The HWs have their own amplitude ($A_i$), phase shift ($P_i$), frequency ($f_i$), wave length ($L_i$) and a unique number of cycle $\left(\frac{n}{L_i}\right)$.

### 2.3 Forecasting
This study used the developed model to forecast the future demand of house water data. The model used was of the following form:

$$Y_t = \mu + \sum_{i=1}^{\frac{n-1}{2}} B_{1i} \sin(2\pi * f_i * t) + \mu + B_{2i} \cos(2\pi * f_i)\, t + \varepsilon_t \tag{3.3.1}$$

## 3. Results
### 3.1 Cox-Stuart Trend Test
Cox and Stuart trend analysis method was subjected to the household water data. The hypotheses tested, at 5% significance level were:

$H_0$: There is no periodic trend in the household water data
$H_A$: There was periodic trend in the household water data
The test results from r software are as indicated in the table below

**Table 1:** Cox-Stuart Trend Test

| Description | Value |
|---|---|
| Cox-Stuart Statistic | 4.8655 |
| P-value | 0.00001141 |
| Significance level(α) | 0.05 |

From table 4 above, it can be deduced that the p-value = 0.00001141 < the significance level (α) = 0.05. Therefore this study can reject the null hypothesis in favor of the alternative hypothesis and conclude that the household water data has a periodic trend.

### 3.2 Sinusoidal modeling
Since the sample size of this study consisted of 97 observations (odd number) the sinusoidal model structure suitable for this study is indicated below.

$$Y_t = \mu + \sum_{i=1}^{n-1/2} B_{1i} \sin(2\pi * f_i * t) + B_{2i} \cos(2\pi * f_i)\, t + \varepsilon_t \tag{4.2.1}$$

The maximum number of the HWs for the observations is $\frac{n-1}{2} = \frac{97-1}{2} = \frac{96}{2} = 48\ HWs$.
Some of the HWs, WLs and frequencies are as indicated in the table below:

**Table 2:** Summary of the HWs, WLs and $f_i$

| HWs (i) | WLs | $f_i$ | HWs | WLs | $f_i$ |
|---|---|---|---|---|---|
| 1 | 97 | 0.010309 | 25 | 3.88 | 0.257732 |
| 2 | 48.5 | 0.020619 | 26 | 3.730769 | 0.268041 |
| 3 | 32.33333 | 0.030928 | 27 | 3.592593 | 0.278351 |
| 4 | 24.25 | 0.041237 | 28 | 3.464286 | 0.28866 |
| 5 | 19.4 | 0.051546 | 29 | 3.344828 | 0.298969 |
| 6 | 16.16667 | 0.061856 | 30 | 3.233333 | 0.309278 |
| 7 | 13.85714 | 0.072165 | 31 | 3.129032 | 0.319588 |
| 8 | 12.125 | 0.082474 | 32 | 3.03125 | 0.329897 |
| 9 | 10.77778 | 0.092784 | 33 | 2.939394 | 0.340206 |
| 10 | 9.7 | 0.103093 | 34 | 2.852941 | 0.350515 |
| 11 | 8.818182 | 0.113402 | 35 | 2.771429 | 0.360825 |
| 12 | 8.083333 | 0.123711 | 36 | 2.694444 | 0.371134 |
| 13 | 7.461538 | 0.134021 | 37 | 2.621622 | 0.381443 |
| 14 | 6.928571 | 0.14433 | 38 | 2.552632 | 0.391753 |
| 15 | 6.466667 | 0.154639 | 39 | 2.487179 | 0.402062 |
| 16 | 6.0625 | 0.164948 | 40 | 2.425 | 0.412371 |
| 17 | 5.705882 | 0.175258 | 41 | 2.365854 | 0.42268 |
| 18 | 5.388889 | 0.185567 | 42 | 2.309524 | 0.43299 |
| 19 | 5.105263 | 0.195876 | 43 | 2.255814 | 0.443299 |
| 20 | 4.85 | 0.206186 | 44 | 2.204545 | 0.453608 |
| 21 | 4.619048 | 0.216495 | 45 | 2.155556 | 0.463918 |
| 22 | 4.409091 | 0.226804 | 46 | 2.108696 | 0.474227 |
| 23 | 4.217391 | 0.237113 | 47 | 2.06383 | 0.484536 |
| 24 | 4.041667 | 0.247423 | 48 | 2.020833 | 0.494845 |

As it can be noted from table above the highest frequency was 0.49 followed by 0.48 and the least 0.01. The longest WL was 97 followed by 48.50 and the least was 2.02 and the least HW was 1 followed by 2 in that order, the highest was 48.
This study further conducted an estimation process of the parameters of the sinusoidal model in (3.13) above using the Fourier series estimation process. Some of the parameters are as indicated below.

$\hat{B}_{11} = -0.07920$, $\hat{B}_{21} = -0.03755$
$\hat{B}_{12} = 0.04490$, $\hat{B}_{22} = 0.052008$
$\hat{B}_{1,33} = 90.28176$, $\hat{B}_{2,33} = 69.06094$
$\hat{B}_{1,94} = -0.00892$, $\hat{B}_{2,95} = 0.00567$
$\hat{\mu} = 3586.94540$

The mean of the collected data is 3586.94540 as indicated in the estimated parameters above. This study took a step further and conducted the F-test to determine the significant HWs so that we could determine the parameters of the sinusoidal model. The F-test is as indicated in the table below:

**Table 3:** F-test for the HWs

| Sources | SS | df | F-test | Decisions |
|---|---|---|---|---|
| 33rd HW | 0.0081 | 2 | 5.8804 | significant |
| Other HWs | 0.0899 | 92 | | |
| 9th HW | 0.0072 | 2 | 2.5905 | insignificant |
| Other HWs | 0.770 | 79 | | |
| 5th HW | 0.0063 | 2 | 3.8110 | insignificant |
| Other HWs | 0.0021 | 81 | | |

From the F-tests, it was only the 33rd HW that was significant, others were insignificant, therefore the parameters of the sinusoidal model was chosen to be $\hat{B}_{1,33} = 90.28176$, $\hat{B}_{2,33} = 69.06094$ and $\hat{\mu} = 3586.94540$ and therefore the suitable sinusoidal model for the GWASCO water supply data would be:

$$\hat{Y}_t = 3586.9454 + 69.06094\cos(0.5236t) + 90.2817\sin(0.5236t)$$

### 3.3 Forecasting

In-sample forecasting was conducted and comparison was made between the forecasted figures and that of the collected data. The plots are indicated in figure 1 below.

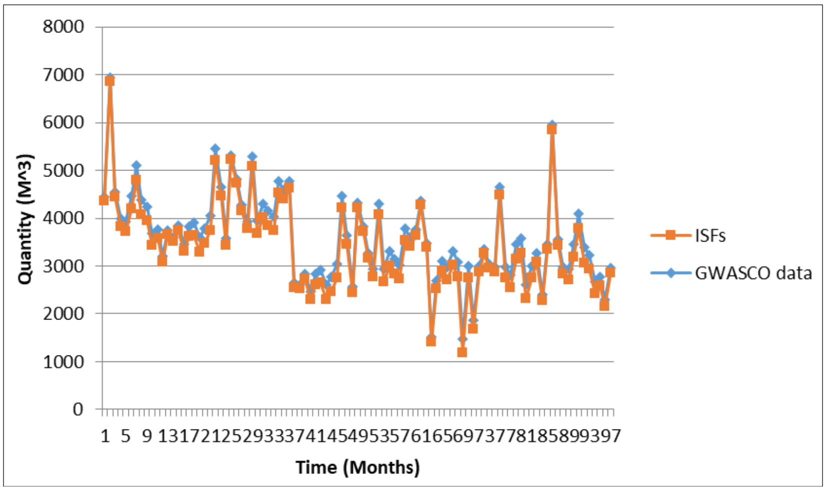


**Fig 1:** In-sample forecasting

From the plot in figure 1 above it can be noted that there was no significant difference in the plots indicating the suitability of the model in forecasting. This study took a step further and conducted the OSFs (future forecasts) 9 months for the next using the developed model. The forecasts are as indicated in the table below.

**Table 3:** Out of Sample Forecasting

| Time (Months) | OSFs |
|---|---|
| 98 | 3697.426 |
| 99 | 3674.984 |
| 100 | 3628.354 |
| 101 | 3570.029 |
| 102 | 3515.639 |
| 103 | 3479.756 |
| 104 | 3471.996 |

| | |
|---|---|
| 105 | 3494.438 |
| 106 | 3541.069 |

## 4. Conclusions

From Cox-Stuart test, this study concluded that the data had a trend and thus could be modeled using the spectral analysis to a data set with a trend.

The spectral analysis was conducted and this study realized that the data had periodic variations and thus the researcher concluded that the data could be modeled using the sinusoids.

In the modeling of the data, this study realized that the number of observations were odd in number (97 months) and thus the researchers estimated and concluded that the number of HWs were 48. The parameters of the model were estimated and the F-test was conducted to determine which of the HWs were significant. It was realized that the $33^{rd}$ HW was significant and thus determined the structure of the model parameters. Thus the parameters were concluded to be $B_1$ = 90.28176 and $B_2$ = 69.06094. The mean of the data was 3586.9454 $M^3$. Thus this study concluded that the model took the following form; $\widehat{Y}_t = 3586.9454 + 69.06094\cos(0.5236t) + 90.2817\sin(0.5236t)$

In sample forecasts was conducted by this study. This study noted no significance difference in the plots of GWASCO water demand data and that of the in-sample forecasts. This study concluded that the fitted model was suitable to be used to describe the data. In effect to the aforementioned, this study used the model to provide 9 months ahead forecast.

## 5. Recommendations

- From this research, we were able to determine the trend of the data using the Cox-Stuart trend test, but the same results could be realized if ordinary least square method in trend determination is employed.
- This study employed spectral analysis of sinusoids in modeling the water demand but future researchers could employ other methods like Box-Jenkins modeling approach and double exponential smoothing methods and make comparison to determine if the same implications could be realized. The model developed in this study could also be investigated to determine if it will be suitable in forecasting the water demand data in other regions of the country.
- Additionally, this study used the in-sample forecasting as a method of determining if the developed model actually represents the true picture of the collected time series data. Future researchers could try other methods like using the residual plots and/ or the coefficient of determination method in determining the suitability of a model.
- Finally, Gusii Water and Sanitation Company (GWASCO) could use the developed model in conducting the future forecasts. This will aid in determining the quantity of water demanded by its customers and thus the company will try to match the amount of water demanded.
- The methods used in this study in modeling the water demand time series data could also be used in other regions by future researchers in modeling the water data.

## 6. Acknowledgement

Special thanks to all the individuals who made the research process a success. We are also indebted to Gusii Water and Sanitation Company (GWASCO) because they availed the household water demand data hence contributing immensely to the success of this research.

## 7. References

1. Mengshan Lee, Berrin Tansel, Maribel Balbin. Influence of residential water use efficiency measures on household water measures on household water demand: A four year longitudinal study. Resources, Conservation and Recycling, 2011, 1-6.
2. Otieno A, Ogembo WO. Domestic water Consumption Per Capita: Acase of Selected Households in Nairobi. Nairobi: University of Nairobi East Africana collection, 2005.
3. Broughton SA, Bryan K. Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing. New york: Wiley, 2008.
4. Berger JSF (Ed.). Bayesian Spectrum Analysis and Parameter Estimation. USA: Springer-Verlag, 1997.
5. Esko Lehtinen *et al*. Statistical Trend Analysis Methods for Temporal Phenomena. SKI Report-Swedish Nuclear Power Inspectorate. 1997; 97(10):732-598.
6. Alan W. Fourier Series (Spectral) Analysis. Personal Teaching, 2004, 1-54.