DISTINCTION BETWEEN SPEPECTRAL ANALYSIS AND HARMONIC ANALYSIS IN TIME SERIES ANALYSIS: EXPERIENCE FROM THE NUMBER CRUNCHER STATISTICAL SYSTEM (NCSS) SOFTWARE.

Authors

Okafor Uchenwa Linus[a], M.O. Oladejo[a], C.O.Uwa[b], D.T.Chineyo[b] and Jighjigh,A.T[c].

[a]. Department of Mathematical Sciences, Faculty of Science Nigerian Defence Academy, Kaduna

[b].Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies ,Nigerian Defence Academy, Kaduna.

[c] Department OF Mathematics/Computer Science. Benue State University, Makurdi

Abstract

This paper shows a clear distinction between Spectral Analysis and Harmonic Analysis and highlights what is common amongst these two terms. The distinction and commonalities are illustrated using the data from sales of cement obtained from a sales record. The work identified the role of spectral analysis as that of showing the periods /wavelengths having containing more energy with the sine and cosine components but without stating whether the components are significant or not. Spectral analysis helps to identify hidden periodicities but does not form the model equation. Harmonic analysis on the other hand makes use of the periods with higher energies identified by the spectral analysis to form a model equation, it also shows whether the frequencies are significant or not by displaying the t-values for each k^{th} *harmonic value used to form the model equation. From the giving data, harmonic analysis makes predictions /forecasts, performs model parameter estimation, Analysis of variance, displays the asymptotic correlation matrix of parameters. In harmonic analysis, the graphs of errors against time and against the predicted values are also plotted.*

Keywords: Harmonic analysis, spectral analysis, period, wavelength, model, forecasting.

Corresponding author's address: email-linusokafor @gmail.com,phone;+234-9022982454

Co-authors address:
(1)mikeoladejo2003@yahoo.com.(2)ozocel2014@gmail.com.(3)atamber@bsum.edu.ng.(4)dtchinyio @yahoo.com

1.0 Introduction

Most text book writers take spectral analysis as harmonic analysis and vice versa.Brillinger(2001) considered harmonic analysis as a tool for search of hidden periodicities. But Rebeca (2004) unlike most of the other authors explained that harmonic analysis works on the assumption that the modeler already knows the period, one can easily set up a regression equation and find out the amplitude, the phase and the mean of the signal. In the analysis of time series data, etc, it is often necessary to identify the periodicities in the data by spectral analysis which is concerned with the partition of the variation in a time series between the components at different frequencies or periods. In spectral analysis, if the data is not stationary it should first be detrended by fitting a proper curve to the data set and then using that to remove the trend. But if the data does not show obvious trend, one can then remove the mean and proceed to carry out spectral analysis. Differencing can also be employed to make the data set stationary. The Number Cruncher Statistical System (NCSS) has six portions on carrying out spectral analysis as shown in Table 1.

Table 1. Fourier Analysis (p,q,s,t,u)

(a)	(b)	(c	(d)	(e)	(f)
Frequencies	Wavelength	Period	Cosine(a_k)	_Sine(b_k)	Spectrum

But we recommend a seven- column portion as given in Table 2. so as to make room for the harmonic number (k)

Table 2. Fourier Analysis (p,q,s,t,u)

(a)	(b)	(c	(d)	(e)	(f)	(g)
Harmonics (k)	Frequencies(f) (f_k)	Wavelength L = $\frac{N}{f}$	Period	Cosine(a_k)	_Sine(b_k)	Spectrum

Where,

$$a_k = \frac{N}{2} * \sum_{k=1}^{\frac{L}{2}} y_k * \text{Cos}(\Theta_k)$$

$$b_k = \frac{N}{2} * \sum_{k=1}^{\frac{L}{2}} y_k * \text{Sin}(\Theta_k)$$

The data points are: Y_k =(y_1 , $y_{2,...}y_n$)

K = $\frac{N}{2}$ for even number of data points, or $\frac{N-1}{2}$ for odd numbered data points.

When the data has a clear seasonal length, L, then

k = $\frac{N}{L}$

The column headed spectrum shows the amount of energy or proportion of the contribution of each of the frequencies to the variance of the series.

Spectral analysis also makes the plots of the spectrum against each of the following items: time, wavelength, and, frequency

It is the spectral analysis that helps one to begin to think of the frequencies /wavelength that will enter into the model building process. Spectral analysis does not give the model, the phase angels and makes no prediction from the data.

In Harmonic analysis, one uses the frequencies identified from spectral analysis to form a model and harmonic analysis gives about eight different specific results as shown through Table 3 to Table 10.Fourier analysis is part of harmonic analysis, but in literature, most authors' mistake spectral analysis for harmonic analysis and most statistical software's also fall into the same misconception, but the use of the Number Cruncher Statistical System(NCSS), illustrates the role of either spectral analysis or harmonic analysis.

2. Statement of the problem.

In order to carry out a Fourier series analysis of any time series data, one must use the periodic components and time series data are seasonal or periodic and more often than not these periodicities are hidden. To unmask the hidden periodic components spectral analysis needs to be performed on the data set. Results from spectral analysis are then used to perform harmonic analysis through Fourier series analysis. There is therefore the need to make a clear distinction between spectral analysis and harmonic analysis in the Fourier analysis of time series data .

3. Aim

The aim of this study is to show the difference between spectral analysis and harmonic analysis in the analysis of time series data through a worked example using the Number Cruncher Statistical System.

4. Data Source and Materials Used.

The data set used in this work is the sales of bags of cement obtained from Ogwumabiri Market in Obowo Local Government Area of Imo State ,Nigeria and this is shown graphically in Figure 1 and the data set is shown in Table 1.The software used to do the analysis of date was the Number Cruncher Statistical System (NCCS2019) by Chris Hintz

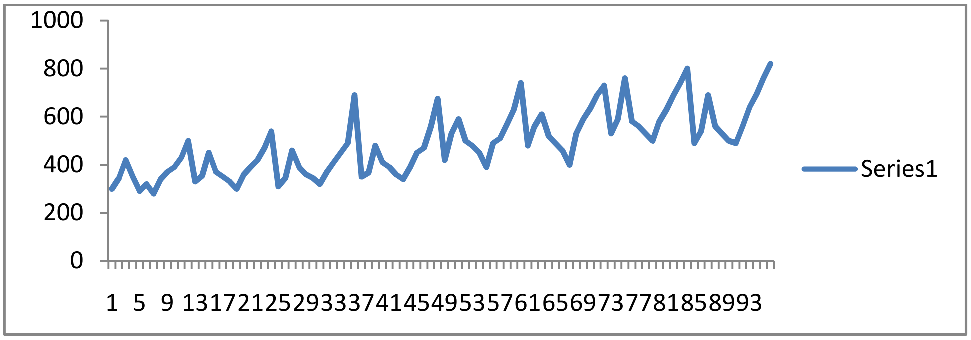

Fig.1.Monthly Sales of Cement 2010 to 2018

Table 1.Sales of Cement 2010 TO 2017 Year

Month/	2010	2011	2012	2013	2014	2015	2016	2017	2018
Jan	300	330	310	350	420	480	530	490	653
Feb	345	355	345	368	530	560	590	540	640
Marc	420	450	460	480	590	610	760	690	670
Apr	350	370	390	410	500	520	580	560	690
May	290	350	360	390	480	490	560	530	610
Jun	320	330	345	360	450	460	530	500	570
Jul	280	300	320	340	390	400	500	490	600
Aug	340	360	370	390	490	530	580	560	640
Sep	370	390	410	450	510	590	630	640	662
Oct	390	420	450	470	570	635	690	695	735
Nov	430	470	490	560	630	690	740	760	790
Dec	500	540	690	675	740	730	800	820	830

5. Methodology

The data was first subjected to outlier test by the Grubbs test for outliers and the (NCSS 2019) found are no outliers.

51. Spectral Analysis

The software (NCSS2019) gave the spectral analysis results in two folds as will be shown in the sequel, it removed the trend as the data displayed in Table 1. depicts trending pattern. Note that if the data has no trend, the mean will be then removed from the data.The spectral plots are displayed in figures 2 and 3.

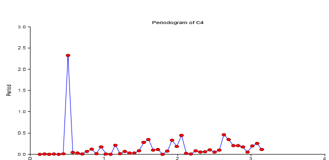

Fig.2 .Plot of Periodogram against Frequency.

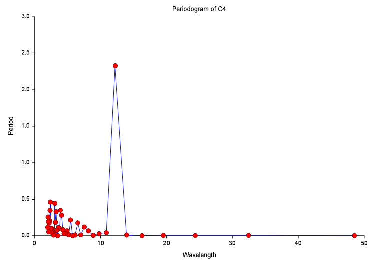

Fig.2.Plot of Periodogram against Wavelength

Table 2.Spectral Analysis Results
Sales of Cement

(k)Harmonic	Frequency	Wavelength	Period	Cosine(a's)	Sine(b's)	Spectrum
1	0.129536	48.50526	16617.3	32.59697	-33.6977	1.00E+33
2	0.193623	32.4507	43948.08	73.84023	18.99913	1.00E+33
3	0.257709	24.38095	29364.56	-53.4418	32.06624	1.00E+33
4	0.321795	19.52542	38437.91	-71.1648	4.476748	1.00E+33
5	0.385881	16.28269	16091.53	-32.3992	-32.8458	1.00E+33
6	0.449968	13.96364	103884.1	-53.0957	-104.51	1.00E+33
7	0.514054	12.22281	2.15E+07	-1541.93	-681.31	1.00E+33
8	0.57814	10.86792	390103.1	206.1583	95.39722	1.00E+33
9	0.642227	9.78344	254522.4	173.7618	58.94443	1.00E+33
10	0.706313	8.895753	60140.59	-88.3929	11.91382	1.00E+33
11	0.770399	8.155752	613223.8	227.667	-171.124	1.00E+33
12	0.834486	7.529412	1110957	-370.889	-96.9354	1.00E+33
13	0.898572	6.992413	116157.1	-103.743	67.841	1.00E+33
14	0.962658	6.526912	1615692	410.1017	213.3926	1.00E+33
15	1.026744	6.119522	84240.29	96.81339	42.07458	1.00E+33
16	1.090831	5.76	11582.41	-29.9363	25.21726	1.00E+33
17	1.154917	5.440378	1979010	-438.565	-263.514	1.00E+33
18	1.219003	5.154362	134286	106.6879	-79.8794	1.00E+33
19	1.28309	4.896918	617032.6	46.99103	-281.8	1.00E+33
20	1.347176	4.663968	290284.8	-186.796	-59.206	1.00E+33
21	1.411262	4.452174	283300.5	181.0477	68.52771	1.00E+33
22	1.475349	4.25878	787854	197.8401	255.0974	1.00E+33
23	1.539435	4.081488	2601302	338.3845	-479.155	1.00E+33
24	1.603521	3.918367	3221233	-652.678	-10.3855	1.00E+33

(k)Harmonic	Frequency	Wavelength	Period	Cosine(a's)	Sine(b's)	Spectrum
25	1.667608	3.767784	919057.6	4.082906	-348.646	1.00E+33
26	1.731694	3.628346	1046506	-294.5	-227.374	1.00E+33
27	1.79578	3.498861	10701.19	37.52555	2.713192	1.00E+33
28	1.859866	3.378299	699923.8	-265.652	148.3689	1.00E+33
29	1.923953	3.265769	3044655	107.2807	625.4841	1.00E+33
30	1.988039	3.160494	1704203	435.7885	-188.458	1.00E+33
31	2.052125	3.061794	4084301	626.947	-383.665	1.00E+33
32	2.116212	2.969072	186455.8	-9.93235	-156.733	1.00E+33
33	2.180298	2.881801	72246.95	-97.085	11.45203	1.00E+33
34	2.244384	2.799514	745561.3	-198.049	-243.716	1.00E+33
35	2.308471	2.721796	512814.4	-169.859	-197.438	1.00E+33
35	2.308471	2.721796	512814.4	-169.859	-197.438	1.00E+33
36	2.372557	2.648276	541663.8	-202.375	175.1979	1.00E+33
37	2.436643	2.578623	951056.5	210.9437	285.1424	1.00E+33
38	2.50073	2.512541	455378.2	147.9449	195.828	1.00E+33
39	2.564816	2.449761	895302.3	-246.868	-239.76	1.00E+33
40	C	2.390042	4245736	422.5156	618.9477	1.00E+33
41	2.692988	2.333164	3190907	-113.368	-639.713	1.00E+33
42	2.757075	2.278932	1882224	-413.898	278.6835	1.00E+33
43	2.821161	2.227163	1877624	441.369	231.4324	1.00E+33
44	2.885247	2.177694	1599362	230.5599	-397.997	1.00E+33
45	2.949334	2.130374	485149.8	211.2075	-139.878	1.00E+33
46	3.01342	2.085068	1806633	-459.92	165.6836	1.00E+33
47	3.077506	2.041648	2362332	154.5963	-537.2	1.00E+33
48	3.141593	2	1065080	-375.348	1.17E-10	1.00E+33

The plots of periodogram against frequency and that of periodogram against wavelengths are shown as Fig.2. and Fig.3 and each of them show that there is only one dominant frequency at the harmonic number k =7 ,with a frequency f= 0.514054 and wavelength 12.22281. .

Table 2 shows the Fourier analysis result indicating the highest energy content at the 7th harmonic k=7 as 2.15*10^{7} ,the cosine component is -1541.93and the sine component is -681.31

7	0.514054	12.22281	2.15E+07	-1541.93	-681.31	1.00E+33

Spectral analysis results end after identification of wavelength, sine term, cosine term, period, and the spectrum. Using the NCSS,the harmonic numbers are not shown, hence we invoke the use of Excel to supplement the analysis in order to add the harmonic numbers as shown in Colum (a) in Table 1.

The use of Harmonic analysis

Harmonic analysis is used to obtain the various section as explained here under.

Below are the steps involved in carrying out the harmonic analysis:

After performing the spectral analysis as discussed above, when the harmonic section of the NCSS2019 was now employed the following results were obtained:

Table 3.Run Summary Section ——

Item	Value	Item	Value
Dependent Variable	Yt	Total Rows	156
Time Variable	t	Rows with Missing Values	60
R^2	0.96	Rows Used	96

Estimated Model

Ft =307.273+(4.317)*t-15.1354)*SIN((0.517*t+ 66.998*COS(0.517t))

The amplitude value of 68.68638 is obtained from the squares of the sine and cosine coefficients 0f;

$$68.68638 = \sqrt{(-15.1354)^2 + (66.998)^2}$$

Table 4.Regression Coefficients Section

Independent Variable	Regression Coefficient b(i)	Standard Error sb(i)	T-Statistic to Test H0: β(i)=0	Prob Level	Upper 95% Conf. Limit of β(i)
Intercept	307	5.59439	54.93	0	318.38
Trend	4.32	0.10026	43.06	0	4.5164
Sin(12.14371)	-15	3.90635	-3.87	0.9999	-7.3771
Cos(12.14371)	67	3.93496	17.03	0	74.813

In this calculation, the software uses the wavelength value of 12.44 ,instead of the the frequency value.

Wave Length	Frequency	Amplitude	Phase
12.144	0.52	68.68638	0.22218

Table 5.Analysis of Variance Table ——

	DF	Sum of Squares	Mean Squares	F-ratio
Intercept	1	25546130	25546130	
Total (Adjusted)	95	1685138		MSE/MSE
Model	4	27163716	MSR=6790928.96	9248.64
Error	92	67552.16	MSE=734.2626	
Total	96	27231268	734.2626	

Table 6.Correlation Matrix of Regression Coefficients

	Intercept	Trend	Sin(12.14371)	Cos(12.14371)
Intercept	1	-0.869221	-0.083666	0.005182
Trend	-0.9	1	0.095877	0.003687
Sin(12.14371)	-0.1	0.095877	1	-0.000566
Cos(12.14371)	0.01	0.003687	-0.000566	1

Table 1.SALES OF CEMENT 2010 TO 2017

Month/Year	2010	2011	2012	2013	2015	2016	2017	2018
Jan	365	420	410	511	630	630	750	735
Feb	345	380	430	525	650	680	640	743
Marc	290	392	400	480	590	700	690	700
Apr	267	380	390	430	520	650	710	670
May	280	365	360	410	530	560	630	660
Jun	275	320	355	420	500	570	610	625
Jul	280	300	340	490	580	550	680	630
Aug	320	360	390	480	530	580	630	640
Sep	350	390	430	520	640	640	680	675
Oct	390	420	490	540	635	690	725	730
Nov	430	470	512	560	690	730	765	760
Dec	450	480	560	590	700	745	780	800

Table 7.Forecasts SALES OF CEMENT 2010 TO 2017

Month/Year	2010	2011	2012	2013	2015	2016	2017	2018
Jan	362	417.4384	472.2456	526.7375	634.72	688	741	741
Feb	337	393.712	450.1702	506.459	618.43	674	729	729
Marc	306	363.18	420.0715	476.9838	590.76	648	704	704
Apr	279	334.9653	390.9591	447.1585	560.08	617	674	674
May	264	317.5843	371.5844	425.921	535.55	591	646	646
Jun	266	316.7174	368.1497	419.9612	524.72	578	631	631
Jul	285	333.7216	382.6842	431.9696	531.56	582	633	633
Aug	318	365.2758	412.5133	459.9327	555.41	604	652	652
Sep	358	404.2496	450.9583	497.6603	591.16	638	685	685
Oct	394	441.5705	489.0851	536.4061	630.57	677	724	724
Nov	419	468.5985	518.0427	567.1571	664.46	713	761	761
Dec	427	479.3884	531.3807	582.9934	685.09	736	786	786

Table 3. forecasts Errors of SALES OF CEMENT 2010 TO 2017

Month/Year	2010	2011	2012	2013	2015	2016	2017	2018
Jan	2.67	2.56165	-62.24556	-15.73755	-4.7183	-58	8.72	-6.28
Feb	7.89	-13.71204	-20.1702	18.54102	31.57	5.93	-89	13.6
Marc	-16	28.82	-20.0715	3.016177	-0.7581	52.4	-14	-4.28
Apr	-12	45.03472	-0.9590976	-17.15847	-40.076	33.3	36.5	-3.51
May	16.1	47.41569	-11.58441	-15.92099	-5.5459	-31	-16	13.7
Jun	9.34	3.282614	-13.14965	0.03875451	-24.72	-7.7	-21	-5.97
Jul	-5.1	-33.72165	-42.6842	58.03038	48.437	-32	47.4	-2.59
Aug	1.81	-5.275815	-22.51331	20.06731	-25.413	-24	-22	-11.9
Sep	-7.5	-14.24964	-20.95833	22.33971	48.843	1.99	-5	-9.96
Oct	-3.8	-21.57052	0.914923	3.593923	4.4325	12.5	0.74	5.74
Nov	11.2	1.401448	-6.042702	-7.157125	25.542	17.3	4.36	-0.64
Dec	23	0.611624	28.61931	7.006566	14.913	9.42	-5.7	14.3

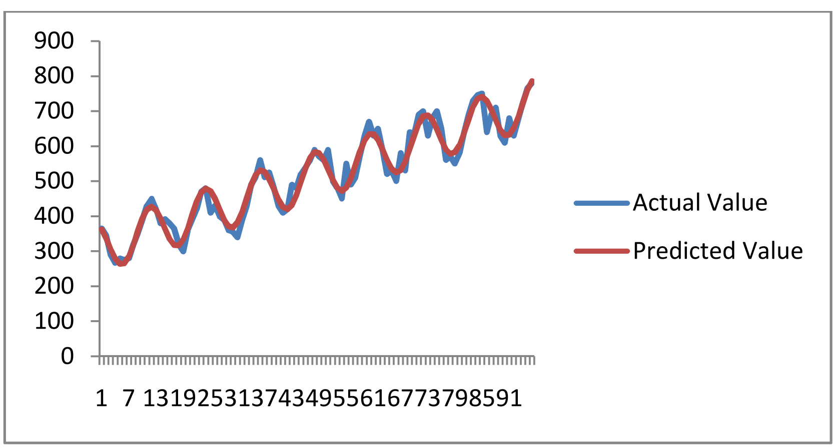

Fig.3.Predicted and Actual values Values against Time

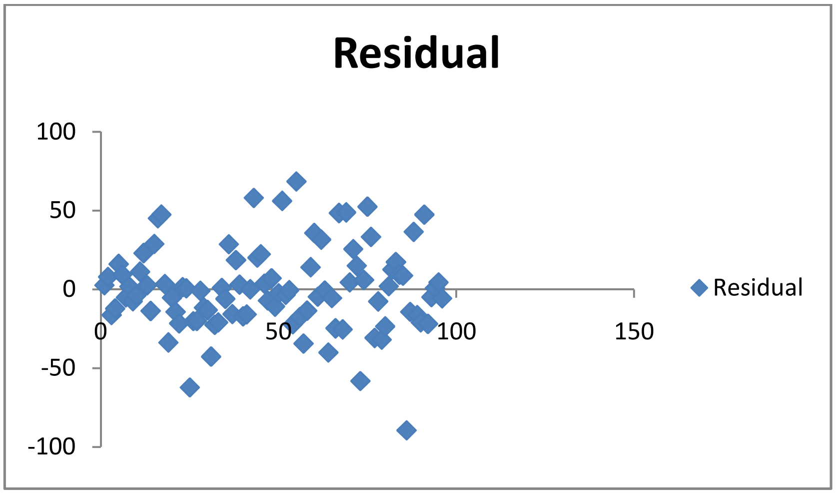

Fig.4.Graph of Residuals against Time

The plot of residuals against time shown in Fig.4 demonstrates that the errors are random in nature, and the plot of actual and predicted values against time shown in Fig.3. is an evidence that that the actual and predicted values are reasonably close to each other.

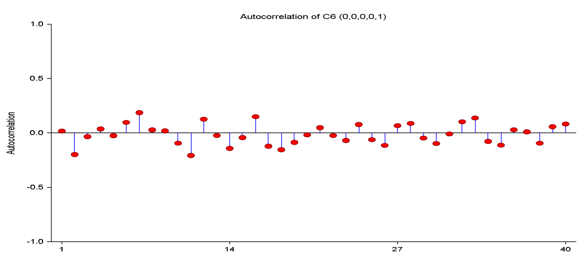

Fig.5.Plot of Autocorrelation Functionsof Errors(ACF).

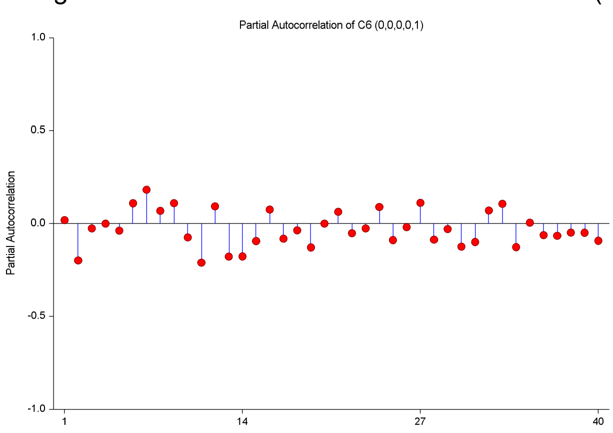

Fig .6. Plot of Partial Autocorrelation Coefficients(PAFC)

Table 4.Autocorrelations of C6 (0,0,0,0,1)

Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.017623	11	-0.206787	21	0.048251	31	-0.009072
2	-0.198659	12	0.126689	22	-0.023265	32	0.103056
3	-0.033041	13	-0.023120	23	-0.068452	33	0.137523
4	0.037400	14	-0.141602	24	0.078515	34	-0.077645
5	-0.024658	15	-0.041887	25	-0.061904	35	-0.112095
6	0.096889	16	0.149367	26	-0.114473	36	0.028800
7	0.186567	17	-0.122003	27	0.067000	37	0.010588
8	0.027944	18	-0.154647	28	0.087605	38	-0.094329
9	0.019432	19	-0.086806	29	-0.046813	39	0.058022
10	-0.094473	20	-0.017866	30	-0.097414	40	0.082354

Significant if |Correlation|> 0.204124

Table 5.Partial Autocorrelations of Yt (0,0,0,0,1)

Lag	Correlation	Lag	Correlation	Lag	Correlation	Lag	Correlation
1	0.017623	11	-0.210698	21	0.062361	31	-0.099679
2	-0.199031	12	0.092510	22	-0.052010	32	0.070056
3	-0.026389	13	-0.178303	23	-0.026525	33	0.106009
4	-0.001082	14	-0.177207	24	0.089137	34	-0.127389
5	-0.038603	15	-0.094584	25	-0.089422	35	0.004287
6	0.108780	16	0.074532	26	-0.020151	36	-0.062159
7	0.181286	17	-0.082308	27	0.110811	37	-0.065611
8	0.067983	18	-0.036400	28	-0.087167	38	-0.049178
9	0.109374	19	-0.129148	29	-0.030008	39	-0.049802
10	-0.075281	20	-0.000654	30	-0.124178	40	-0.092858

Significant if |Correlation|> 0.204124

The results from both Tables 5 and 6 show that that the PAC F,and ACF have absolutes values that are less than the table value of 0.204124 as indicated by the software.This further confirms the adequacy of the adopted model.The plots of ACF and PACF shown in Figures 4 and 5 each did not exhibit spike at any lag,demonstrating that the correct model has been obtained.

Conclusion.

The paper has explained briefly the difference between spectral analysis and harmonic analysis given the results in each case using an example. It has also shown that to carry out harmonic analysis, spectral analysis must be performed to obtain significant frequency that will be used to get an appropriate model.

References

1.David R. Brillinger ((2 NumberZ001) Time series Data Analysis and Theory.Holden Day.Inc,San Francisco.

2.Chris Hintz(2019) .number Cruncher Statistical System (NCSS 2019)

3.