CARD International Journal of Science and Advanced Innovative Research (IJSAIR)
ISSN: 2536-7315 (Print) 2536-7323 (Online)
Volume 2, Number 3, September 2017
http://www.casirmediapublishing.com

A Comparison of the Pseudo-Additive Mixed Fourier series Approach and the SARIMA Methodology to the Modeling of Rainfall in Uyo

Anyanso Chinomso
Department of Mathematics/ Statistics
Faculty of Science, Rivers State University, Port Harcourt

ABSTRACT

Accurate rainfall forecasting is very important to the economic development of a country. It is not just important to the government but also to individuals, farmers and private companies. This paper focuses on comparing the performances of two approaches to seasonal time series analysis. These approaches are the pseudo-additive mixed Fourier series approach and the SARIMA approach. The pseudo-additive Fourier series approach decomposes a time series into the traditional components in a mixed model. This is suitable for a time series with very small or zero values like that in the data used, while the ARIMA model has significant advantages especially in short run forecasting saz [2011], The time series analysis methods were used to model the monthly rainfall of Uyo in Akwa Ibom state, Nigeria . The data were monthly value for ten[10] years .A comprehensive outline of both analysis methods are presented in this paper as well as the advantages each have after the other . The performances were evaluated based on three[3] statistics; mean absolute error [MAE], mean absolute percentage error[MAPE] and mean squared deviation [MSD], The result at the end showed that the SARIMA model has a smaller MAE, MAPE and MSD values .As such, it is the better model.

INTRODUCTION

Rainfall is a very important feature of our climate and its importance (economically and otherwise) cannot be overemphasized. For an economy like ours where diversification into Agriculture is being encouraged, rainfall is a very important element which cannot be overlooked. Other sectors which are also vulnerable to weather variability include tourism, mining, construction etc.

One of the characteristics of rainfall in Nigeria is seasonality. The rainy season is experienced from March till October, with peak rainfall in August. The dry season follows from late October till early March. Sometimes due to too much rainfall, soil moisture reaches saturation levels and the soil is no longer able to absorb more rainwater. This leads to over flooding which coincides with the peak of the rainy season. In 2012, there was flooding in Nigeria. According to a report on Punchng.com (2015), the total value of damaged physical and durable assets caused by floods in the most affected states was estimated to have reached #1.45trillion. There is also the issue of drought. It is on record that the 1972/73 drought drastically reduced the contribution of agriculture to the Gross Domestic Product in Nigeria from 18.4% in 1971/72 to 7.3% in 1972/73.These and other issues have made the forecasting of rainfall to be of great importance.

Some models have been applied to various time series on rainfall, chief of which is the Seasonal Autoregressive Integrated Moving Average Model (SARIMA). However, this paper seeks to compare the popular SARIMA approach to a frequency domain based approach using linear regression and Fourier series analysis methods. Both approaches shall be used to model the available data and the results of the estimated values compared using relevant statistics.

LITERATURE REVIEW

Traditional time series analysis methods involved the identification, decomposition and estimation of the basic components. Etuk and Uchendu (2009) listed the basic components of a time series as the trend, seasonal component, cyclical component and irregular component. In a draft guide to seasonal adjustment with X-1-ARIMA, the United States Office of National Statistics, Time Series Analysis Branch (2007), discussed the decomposition models and outlined them mathematically as;

a. $X_t = T_t+S_t+I_t$
(Additive model)

b. $X_t = T_t\times S_t \times I_t$
(Multiplicative Model)

In practice, most economic time series exhibit a multiplicative relationship and hence, the multiplicative decomposition usually provides the best fit. However this decomposition approach cannot be implemented if any zero or negative values appear in the time series. This is because it is not possible to divide a number by zero. This leads to another decomposition model.

c. $X_t = T_t(S_t + I_t + 1)$
(Pseudo additive model)

Where,

T_t= linear trend

S_t = seasonal component

I_t = irregular component

For a time series that contains very small or zero values like that which is used for this work, the Pseudo-additive model is used. In this model, both the seasonal and irregular component are centered around one. Etuk and Uchendu (2009) states that the trend can be calculated using the method of ordinary least squares. The seasonal component is based on the work of Joseph Fourier who introduced the series

$$x_t = \sum_{j=0}^{n} (\alpha_j Cos(w_j t) + \beta_j Sin(w_j t))$$

The equation above seeks to represent a periodic series as a sum of sinusoidal components. Lewis (2003) applied Fourier Analysis to the forecasting the inbound call time series of a call centre. He found out that working in the frequency domain overcomes many difficulties encountered in the time domain. Omekara *et al.*, (2013) modeled the Nigerian inflation rates using Fourier series and periodogram and disclosed that it is better than the time domain approach because of their simple way of modeling seasonality and eliminating peaks without re-estimating the model.

Using the SARIMA methodology, a Time series with span of seasonality s, is said to follow a multiplicative (p, d, q) × (P, D, Q)$_s$ seasonal ARIMA model if

$A(L)\Phi(L^s)\nabla^d\nabla_s^d X_t = B(L)\Theta(L^s)e_t$

A(L) and B(L) refers to the non-seasonal AR component and MA components respectively. **$\Phi(L^s)$ and $\Theta(L^s)$** are the respective seasonal AR and MA component. e_t is the error term and ∇^d ***and*** ∇_s^d are the regular and seasonal differencing respectively required to make the Time series stationary. A time series is said to be stationary if it has a constant mean, a constant variance and an autocorrelation that is a function of the lag separating the correlated values (Etuk, 2013). Most real-life Time series are non-stationary. And since stationarity is a requirement for using this approach, Box and Jenkins proposed that such a series could be made stationary by differencing with an appropriate order.

A very useful operator is the backward linear operator (B) or lag operator defined by:

$BX_t = X_{t-1}$

$B^2X_t = BBX_t = BX_{t-1} = X_{t-2}$

$B^3X_t = BB^2X_t = BX_{t-2} = X_{t-3}$

$B^kX_t = X_{t-k}$

Since the mean is constant throughout the series,

$B\mu = \mu$

In order to make a non-stationary series to become stationary, we apply the difference operator (**∇**), which is defined by

$$\nabla = 1 - B$$

$\Rightarrow \nabla X_t = (1-B)X_t = X_t - X_{t-1}$

$\nabla^2 X_t = (1-B)^2 X_t = (1-2B+B^2)X_t$

$= X_t - 2X_{t-1} + X_{t-2}$

The stationary series W_t obtained as the dth difference (**∇^d**) of X_t is given by

$W_t = \nabla^d X_t = (1-B)^d X_t$

For a seasonal Time series, a seasonal differencing is first carried out and then a regular differencing, if still required.

METHODOLOGY

The data for this work are monthly rainfall data for the city of Uyo in Nigeria for ten (10) years. This data was used in a research published in the Asian Journal of Mathematics and Statistics (ISSN 1994-5418).

Using the frequency domain approach first, recall that the model is given by

$$X_t = T_t(S_t + 1)$$

The trend is obtained by using the method of ordinary least square

i.e $X_t = a_o + b_o t$

After the trend is obtained, the series is detrended and the seasonal component is estimated thus:

$$DT = \frac{Xt}{Tt} = (S_t + 1)$$

And the seasonal component is

$$St = DT - 1$$

The model for seasonality is

$$St = \sum_{i=1}^{k}(a_i \cos \omega it + b_i \sin \omega it).. \qquad (1)$$

$Where, k = number\ of\ observation\ per\ season\ divided\ by\ 2$

i.e. $k = \frac{12}{2} = 6$

$\omega = \frac{2 \times \pi \times f}{n} = \frac{2 \times \pi \times 10}{120} = \frac{\pi}{6}$

Substituting the above in equation 1, gives

$$S_t = \sum_{i=1}^{6}\left(a_i \cos\frac{i\pi t}{6} + b_i \sin\frac{i\pi t}{6}\right). \qquad (2)$$

The above equation is cast as a multiple linear regression to obtain the estimates of a_i and b_i. Finally, we obtained the residuals

$E_t = X_t - \tilde{X}_t$

This is tested for randomness by checking the autocorrelation function to see if there are significant autocorrelations.

For the SARIMA methodology, the analysis starts with a time plot. The plot is examined for any of the traditional components such as trend, seasonality etc. The presence trend and/or seasonality signify non-stationarity. Where there is seasonality, a seasonal differencing is first carried out. If the series is still not stationary, then a regular differencing is needed. An augmented Dickey-Fuller test can be used to test for stationarity. In order to identify a suitable model, a look at the graph of the seasonally-differenced autocorrelation function (ACF) is necessary. A significant spike at the seasonal lag gives a clue of the model. If the spike is positive, it suggests a seasonal autoregressive model and if it is negative, it is a seasonal moving average model. In addition, a spike in the early lags of the ACF and PACF is indicative of the MA and AR components respectively. The cut-off point on the correlogram gives the order. All this can be done using the statistical software, E-views. It applies the least squares approach to estimation. A SARIMA $(p, d, q) \times (P, D, Q)_s$ model is a seasonal Autoregressive Integrated Moving Average Model with non-seasonal AR and MA components of orders p and q respectively. The number of regular differencing required is d. In addition, there are seasonal AR and MA components of orders P and Q respectively. D is the number of seasonal differencing carried out and s is the span of seasonality. Often the value of d and D does not exceed 1.

After fitting a model, it must be subjected to residual analysis to ascertain that the residuals are uncorrelated. If the residuals are correlated, a better model has to be used.

RESULTS AND DISCUSSION

A look at the time plot of the original data shows seasonality. Since the data is monthly, the span of seasonality is 12. From the trend analysis,

$a_o = 210.4116$

$b_o = -0.059$

The MS-excel output is given below

Table 1: Results for the linear trend analysis

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	210.4116	29.10406	7.229631	5.22E-11	152.7776	268.0456	152.7776	268.0456
X Variable 1	-0.05974	0.417473	-0.14309	0.886461	-0.88645	0.766972	-0.88645	0.766972

From the above, b_o is clearly not significant. Hence T_t = 210.4116

The seasonal component is modeled using equation (2) above. The multiple regression was also done using excel and the output is given in table 4.2

Table 4.2: Result for testing the significance of the detrended series

ANOVA

	df	SS	MS	F	ignificance F
Regression	12	66.8368	5.569733	2657.706	1.7E-126
Residual	108	0.226335	0.002096		
Total	120	67.06313			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Ipper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
X Variable	-0.9828	0.00591	-166.293	5.5E-132	-0.99452	-0.97109	-0.99452	-0.97109
X Variable	-0.21788	0.005913	-36.8479	5.47E-63	-0.2296	-0.20616	-0.2296	-0.20616
X Variable	-0.15307	0.005924	-25.8375	4.8E-48	-0.16481	-0.14133	-0.16481	-0.14133
X Variable	0.099269	0.00595	16.68319	1.16E-31	0.087475	0.111063	0.087475	0.111063
X Variable	0.096637	0.00591	16.35135	5.5E-31	0.084922	0.108351	0.084922	0.108351
X Variable	0.005813	0.005937	0.979088	0.329725	-0.00596	0.017581	-0.00596	0.017581
X Variable	0.194711	0.005926	32.85866	4.74E-58	0.182965	0.206456	0.182965	0.206456
X Variable	0.015869	0.005914	2.683469	0.008433	0.004147	0.02759	0.004147	0.02759
X Variable	-0.13081	0.005918	-22.1029	7.03E-42	-0.14254	-0.11908	-0.14254	-0.11908
X Variable	-0.02645	0.005922	-4.46704	1.96E-05	-0.03819	-0.01471	-0.03819	-0.01471
X Variable	0.029093	0.004851	5.997716	2.7E-08	0.019478	0.038708	0.019478	0.038708
X Variable	-3.6E+11	3.09E+11	-1.17919	0.240912	-9.8E+11	2.48E+11	-9.8E+11	2.48E+11

From the above, the equation for the seasonal component is given as;

S_t = -0,9828cosωt – 0.21788sinωt – 0.15307cos2ωt + 0.099269sin2ωt +0.096637cos3ωt + 0.005813sin3ωt + 0.194711cos4ωt – 0.13081cos5ωt – 0.02645sin5ωt + 0.029093cos6ωt

The fitted values obtained using this procedure is given in appendix 2. The auto correlogram of the residuals show that they are uncorrelated.

For comparative purposes with the SARIMA approach, the following statistics were calculated:

MAE = $\sum_{i=1}^{n} \frac{/Yi-\tilde{Y}i/}{n}$ = 5.347117

MAPE = 100 $\sum_{i=1}^{\wedge} \frac{/yi - \hat{y}_i /}{yi}$ = 5922.968

MSD = $\frac{1}{n}\sum_{i=1}^{n}(Yi - \tilde{Y}i)^2$ = 84.83943

The time plot of the original series shows an upward movement from the third month which peaks usually at the eighth month and then a downward movement of the values follow. This is indicative of seasonality. Since this occurs every 12 months, it is seasonality of span 12. As a result, a seasonal differencing of span 12 was carried out. The time plot of the seasonally differenced series suggests stationarity. An ADF test proves that the seasonally differenced series is stationary.

Thus there is no need for regular differencing any longer. The correlogram of SDUYOR shows a negative spike at lag 12 which is indicative of a seasonal MA of order 1. No other component is visible and thus the model is SARIMA $(0,0,0)\times(0,1,1)_{12}$. The equation for this model is given by

$Y_t = Y_{t-12} + \Theta_1 e_{t-12} + e_t$

Using MINITAB, the output after applying this model is shown in the table below.

Table 4.3: Final Estimates of Parameters

Type	Coef	SECoef	T	P
SMA 12	0.8864	0.0682	12.99	0.000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 120, after differencing 108

Residuals: SS = 8971.29 (backforecasts excluded)

MS = 83.84

DF = 107

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	8.5	17.4	30.3	37.8
DF	11	23	35	47
P-Value	0.665	0.790	0.696	0.829

Substituting the value for the seasonal MA component, the equation becomes

$\tilde{Y}_t = Y_{t-12} + 0.8864 e_{t-12} + e_t$

The fitted values are given in appendix 3 and the model is seen to closely agree with the data. The correlogram of the residuals shows that they are uncorrelated.

Also, for comparative purpose,

MAE = 4.74241

MAPE = 4016

MSD = 83.0675

CONCLUSION

It can be inferred from the above that the Pseudo-additive Fourier series model and the SARIMA model are quite suitable for the rainfall data. As such, they can be used to model and forecast a seasonal time series. In addition, comparing the MAPE, MAE and MSD obtained from the output from both models, it can be seen that the SARIMA model outperforms the Pseudo-additive model.

APPENDIX 1:
ACTUAL VALUES

S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y
1	6.2	13	4.5	25	6.2	37	0	49	8.8	61	2.9	73	5.3	85	5.7	97	4.8	109	0.2
2	56	14	60.5	26	55	38	45	50	50.9	62	53.8	74	55	86	57	98	45	110	42
3	200.4	15	221	27	219	39	222	51	224	63	225	75	225	87	222	99	230	111	217
4	290	16	299	28	300	40	297	52	310	64	296	76	299	88	297	100	287	112	290
5	273	17	271	29	280	41	277	53	268	65	273	77	273	89	272	101	277	113	268
6	450	18	439	30	438	42	440	54	440	66	400	78	433	90	436	102	444	114	442
7	371	19	372	31	372	43	372	55	382	67	372	79	372	91	300	103	356	115	372
8	396	20	400	32	397	44	397	56	396	68	395	80	401	92	389	104	389	116	403
9	320.5	21	331	33	323	45	328	57	325	69	324	81	328	93	323	105	333	117	321
10	89.1	22	90.1	34	87.4	46	88	58	100	70	88	82	92.6	94	90.3	106	92	118	97
11	10.9	23	15	35	14	47	12	59	15	71	9	83	10	95	11.6	107	12	119	13
12	14	24	10	36	9.9	48	10	60	3.5	72	11.2	84	5.8	96	12.5	108	10	120	1.1

APPENDIX 2:
ESTIMATES USING FOURIER SERIES APPROACH

S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y
1	9.131	13	9.29	25	9.18	37	9.07	49	7.88	61	8.84	73	5.5	85	8.62	97	7.4	109	6.2
2	55.15	14	55.3	26	54.8	38	56	50	55.1	62	56.3	74	55.3	86	56.5	98	58	110	55
3	224.8	15	224	27	224	39	223	51	224	63	225	75	224	87	223	99	226	111	225
4	299.5	16	300	28	300	40	300	52	299	64	300	76	301	88	302	100	299	112	301
5	277.4	17	278	29	277	41	276	53	277	65	276	77	277	89	278	101	277	113	276
6	439.1	18	439	30	440	42	439	54	440	66	437	78	442	90	439	102	441	114	442
7	368.2	19	368	31	368	43	368	55	369	67	367	79	368	91	367	103	366	115	369
8	399.6	20	400	32	400	44	401	56	400	68	399	80	400	92	401	104	398	116	400
9	329.9	21	330	33	331	45	330	57	329	69	330	81	328	93	327	105	330	117	329
10	94.56	22	95.2	34	94.2	46	95.4	58	94.5	70	95.7	82	96.8	94	93.7	106	95	118	96
11	15.77	23	15.7	35	15.5	47	16.5	59	15.3	71	18.4	83	17.3	95	16.1	107	15	119	14
12	11.63	24	11.7	36	11.9	48	12	60	13.2	72	12.2	84	11.2	96	12.4	108	14	120	15

APPENDIX 3:
ESTIMATES USING THE SARIMA METHODOLOGY

S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y	S/No	Y
1		13	5.05	25	4.99	37	5.13	49	4.55	61	5.03	73	4.79	85	4.85	97	4.9	109	4.9
2		14	51.4	26	52.4	38	52.7	50	51.8	62	51.7	74	52	86	52.3	98	53	110	52
3		15	220	27	220	39	220	51	220	63	221	75	221	87	222	99	222	111	223
4		16	293	28	294	40	295	52	295	64	297	76	297	88	297	100	297	112	296
5		17	275	29	274	41	275	53	275	65	274	77	274	89	274	101	274	113	274
6		18	441	30	440	42	440	54	440	66	440	78	435	90	435	102	435	114	436
7		19	364	31	365	43	366	55	366	67	368	79	369	91	369	103	361	115	361
8		20	394	32	395	44	395	56	395	68	395	80	395	92	396	104	395	116	394
9		21	327	33	328	45	327	57	327	69	327	81	327	93	327	105	326	117	327
10		22	90.8	34	90.7	46	90.4	58	90.1	70	91.2	82	90.9	94	91.1	106	91	118	91
11		23	12.2	35	12.5	47	12.7	59	12.6	71	12.9	83	12.5	95	12.2	107	12	119	12
12		24	10.4	36	10.4	48	10.3	60	10.3	72	9.51	84	9.7	96	9.26	108	9.6	120	9.7

REFERENCES

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). Time Series Analysis, Forecasting and Control. (4th edition), John Wiley and Sons, New Jersey.

Omekara, C. O., Ekpeyong, E. J. &Ekerete, M. P. (2013). Modeling the Nigeria Inflation Rates Using Periodogram and Fourier Series Analysis. *CBN Journal of Applied Statistics* vol. 4 No. 2 Pp 51 – 68.

Emmanuel, J. E. (2008). Pseudo-Additive (mixed) Fourier Series Model of Time Series. *Asian Journal of Mathematics and Statistics*

Etuk, E. H. (2013). Multiplicative SARIMA Modeling of Daily Naira-Euro Exchange Rates. *International Journal of Mathematics and Statistics Studies*Vol 1. No. 3 Pp 1 – 8.

Etuk, E. H. &Uchendu, B. A. (2009), Basic Statistics With Applications, Vol. II, Owerri, Duke Press

Saz, G. (2011). The Efficiency of SARIMA model for forecasting Inflation Rates in Developing countries: The case for Turkey: *International Research Journal and Finance Economics*, (62), Pp 111 – 142.

ONS (2007). Guide to Seasonal Adjustment with X-12-ARIMA. National Bureau of Statistics

Punchng.com, (2005). NIMETS 2016 seasonal Rainfall Prediction. Retrieved 08 September, 2016, from http://punchng.com/nimets-2016-seasonal-rainfall-prediction.