

ON THE MODELLING OF ROAD TRAFFIC CRASHES: A CASE OF SARIMA MODELS

Eze, C. M.^{1*}, Asogwa, O. C.², & Okonkwo, C. R.³

¹Department of Statistics, University of Nigeria Nsukka Tel: +2348069464222 E-mail: chinonso.eze@unn.edu.ng

²Federal University Ndufu Alike Ikwo, Ebonyi State of Statistics Tel: +2348036785840; E-mail: qackasoo@yahoo.com

³Okonkwo, Chinonye Rosemary Department of Statistics, University of Nigeria Nsukka Tel: +2348038704445 E-mail: steveokoyeis@yahoo.com

***Corresponding Author: -**

E-mail: chinonso.eze@unn.edu.ng

Abstract: -

This paper examined the modeling of accident cases in four major roads leading to the main city of Enugu State of Nigeria using SARIMA Models. Among the most robust approaches for analysing time series data is the Autoregressive Integrated Moving Average (ARIMA) model propounded by Box and Jenkins (1979). In this paper, we employed the Box-Jenkins methodology to build SARIMA model for the accident cases for the period, January 2007 to December 2015 with a total of 108 data points. The model obtained in this paper was used to forecast monthly cases of accident in each of the roads for the upcoming year 2016. The forecasted results will help Government and Federal road safety commission to see how to maintain orderliness on the roads to reduce the case of road traffic crashes along the roads

.



Distributed under Creative Commons CC BY-NC 4.0 OPEN ACCESS

1.1 BACKGROUND OF STUDY

Before the advent of modern transportation system, Human beings and goods (wares) moved from one place to another either by trekking or by the use of animals like donkeys, camels and horses. Today, technology has birthed faster means of transportation from one place to another. It has been posited that a good means of transportation serves as a lifeline of an economy (Olubomehin, 2012). Indeed, the provision of a high-quality transport system has been identified as a pre-condition for the full participation of remote communities in the benefit of natural development. Among the various means of transportation known to man, road transportation is the most common and the most important since it is relatively more affordable and accessible by people. Thus, it is exposed to more mishaps than other means of transportation. Road traffic crashes (RTC) has been noted to be the leading cause of death in adolescents and adults worldwide. Currently, it ranked 9th position in order of disease burden and are projected to be ranked 3rd in the year 2020 (Atubi, 2012). Avuglah et al, (2014) in their Research have indicated that about 50 million people are injured and 1.2 million people are killed each year in road accident worldwide. Traffic crashes and injuries have incurred an annual loss of \$65 billion to \$100 billion (Agbeboh and Osabuohien, 2013). In Nigeria, it has been estimated that persons injured in accidents on the highways and streets no longer participate in the economic mainstream and this amounts to a loss of labour of millions of persons in the nation (Pratte, 1998).

Accident is an event, occurring suddenly, unexpectedly and inadvertently under unforeseen circumstances. Road traffic crash/accident can be said to be an unplanned occurrence of auto crash that may result in injuries, loss of lives and properties. It claims the largest toll of human life and tends to be the most serious problem all over the world. Although transportation has liberated man and makes him more mobile, yet his increasing reliance on vehicular movement has conferred great fatalities on him and his activities (Sumaila, 2013).

Ironically, in Nigeria, studies have indicated that better facilities in terms of good quality and standardized roads have been accompanied by increasing number of accidents (Atubi and onakala 2009). This is totally contrary to the trends in countries where even the level of sophisticated road network and volume of vehicular traffic are much higher. In an effort to check this alarming trend, the Nigerian Federal government saw the need to establish the present federal Road Safety Commission in 1988 to address the road safety carnage on the high way. This commission has helped through their operations in minimizing Road traffic crash but has not succeeded in eradicating it entirely.

The cases of accident being a timely occurrence can be modeled using Box and Jenkins approach to time series modeling developed by two mathematicians, George Box and Gwilym Jenkins (1970).

The approach tries to identify the appropriate other of a linear model by examining the correlogram of the series. It is a mathematical model designed to forecast data within a time series. The Box- Jenkins model alters the time series to make it stationary by using the differences between data points. This allows the model to pick out trends, typically using autoregression, moving averages and seasonal differencing in the calculations. Autoregressive integrated moving average (ARIMA) model is a form of Box-Jenkins model. Our interest in this study therefore, is to identify a model from ARIMA family that understudies the generative pattern of the cases of accident in each of the routes (Enugu-Abakilike, Nsukka-9th mile, Enugu-Onitsha, Enugu-Porthacourt) and use the model for prediction into the future. When seasonality is contained in the series, the seasonal components are incorporated into the ARIMA model to make the seasonality not to die out. This leads to an extended model known as seasonal autoregressive moving average (SARIMA) model.

Many literatures abound in the area of road traffic crash cases in Nigeria and beyond. Some of the studies (Atubi (2013), Oppong (2012), Olushina et al (2011) and olujimi (2016)) concentrated on the fatality (death) caused by road traffic crashes while others (Salako et al (2014), Iwok (2016), Sumaila (2013) etc) tries to fit a model that could be used to make prediction (forecast) into the future of accident cases. It has been identified that some variables could trigger crashes on the road. This was buttressed in the work of Atubi (2012), where he regressed road traffic crash on the length of roads, presence of road safety and population. It was found that the effects of these selected independent variables are quite significant. The presence of seasonal effects cannot be ignored in time series modeling of accident cases as seen in the work of Katleho (2016). However, Agbeboh and Osabuohien (2013) fitted only the linear equation with the claim that no seasonal component was detected in the cases of accident in Kogi State Nigeria.

Materials and methods

As pointed out earlier, the time series modeling approach, Autoregressive Integrated Moving Average (ARIMA) models propounded by Box and Jenkins (1970) was adopted in this work. The methodology is a modeling technique that expresses a series as a function of an Autoregressive (AR) process, where there is a memory of past values and a Moving Average (MA) process, which accounts for previous error terms, inculcating an Integrated (I) process so as to stabilize the data.

ARIMA (Autoregressive Integrated Moving Average) Models

This is a statistical analysis model that uses time series data to predict future trends. This model type is generally referred to as ARIMA (p, d, q), with the integers referring to the autoregressive, integrated and moving average parts of the data set, respectively. The values of p and q are the number of autoregressive (AR) and moving average (MA) components in ARIMA (p, d, q) model. These two simple components are used in representing the behavior of observed time series process. The AR is used to describe a time series in which the current observation depends on its preceding values, where as the moving average (MA) is used to describe a time series process as a linear function of current and previous random errors. It is possible that a time series model will consist of a mixture of AR and MA components. In this case the series is said to be an autoregressive moving average process of order (p, q), where p and q are the orders of the AR and MA components respectively. The selection strategy for such models was developed and selected by the Box-Jenkins method.

ARIMA Model consists of a three stage iterative process, which comprises of identification of the model, parameter estimation and diagnostic checking of residuals of the fitted model. Tentative identification of ARIMA process is done through analysis of actual historical data. In general, we should have atleast 50 observations available to identify the appropriate model satisfactorily. Tentative identification is usually done by observing the behavior of the theoretical autocorrelation and partial autocorrelation functions. Model estimation means finding the values of the model coefficients which provide the best fit to the data. Let the time series under consideration be Z_t . If d is a non-negative integer, then the $\{Z_t\}$ is said to be an ARIMA (p, d, q) process if $Y_t =$

$\nabla^d Z_t$ is a causal autoregressive moving average (ARMA) process. Z_t satisfies a difference equation of the form: $\phi(B)\nabla^d Z_t = \theta(B)\varepsilon_t$. Where B is the backward shift operator, $\nabla = 1 - B$, $B^m Z_t = Z_{t-m}$, $\{\varepsilon_t\} \sim N(0, \sigma^2)$, $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q respectively, $\phi(B) \neq 0$ for $|B| \leq 1$. the polynomial $\phi(B)$ has a zero of order d at $B = 1$. The process $\{Z_t\}$ is said to be stationary if $d = 0$, in which case the ARIMA (p, d, q) reduces to an ARMA (p, q). The ARIMA model is a very useful statistical method for analyzing data with a correlation among neighbouring observations.

SARIMA (Seasonal Autoregressive Integrated Moving Average) Models

This is an extension of ARIMA models. Seasonal ARIMA (p, d, q) is used when seasonal (hence non stationary) behavior is present in the time series. The Seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model with the form SARIMA (p,d,q)(P,D,Q)_S, where:

- p, d, q are the parameters in non-seasonal ARIMA model.
- P is the number of seasonal Autoregressive order,
- D is the number of seasonal differencing,
- Q is the number of seasonal Moving Average order, and
- S is the time span of repeating seasonal pattern.

In a mathematical expression, the SARIMA model can be presented as $\phi_p(B)\phi_p(B^S)\nabla^d\nabla^D Z_t = \theta_q(B)\theta_q(B^S)a_t$. $\phi_p(B)$ is the transfer function for the non-seasonal autoregressive factor, $\phi_p(B^S)$ is the transfer function for the seasonal autoregressive factor, ∇^d is the non-seasonal differencing of the series of order, d , ∇^D is the seasonal differencing of the series of order, D , Z_t is the time series observations, $\theta_q(B)$ is the transfer function for the non-seasonal moving average factor, $\theta_q(B^S)$ is the transfer function for the seasonal moving average factor and a_t is the error at time, t . The ability to model complex seasonal time series greatly increases the applicability and usefulness of autoregressive integrated moving average (ARIMA) model building proposed by Box and Jenkins (1970). Many time series such as climate, economic, accident etc. are observed to exhibit some periodic and recurrent nature. In such cases, the popular ARIMA models cannot provide good approximations for the true underlying process. Hence the need for a component in ARIMA model that caters for the periodic influences called seasonal ARIMA (SARIMA) model. However, in practice, it may not be reasonable to assume that the seasonality component repeats itself precisely in the same way cycle after cycle (Brockwell, 1986). Seasonal ARIMA models do allow for randomness in the seasonal pattern from one cycle to the next. In general, we say a series exhibits periodic behavior with period S , when similarities in the series occur after S basic intervals.

Data analysis and results

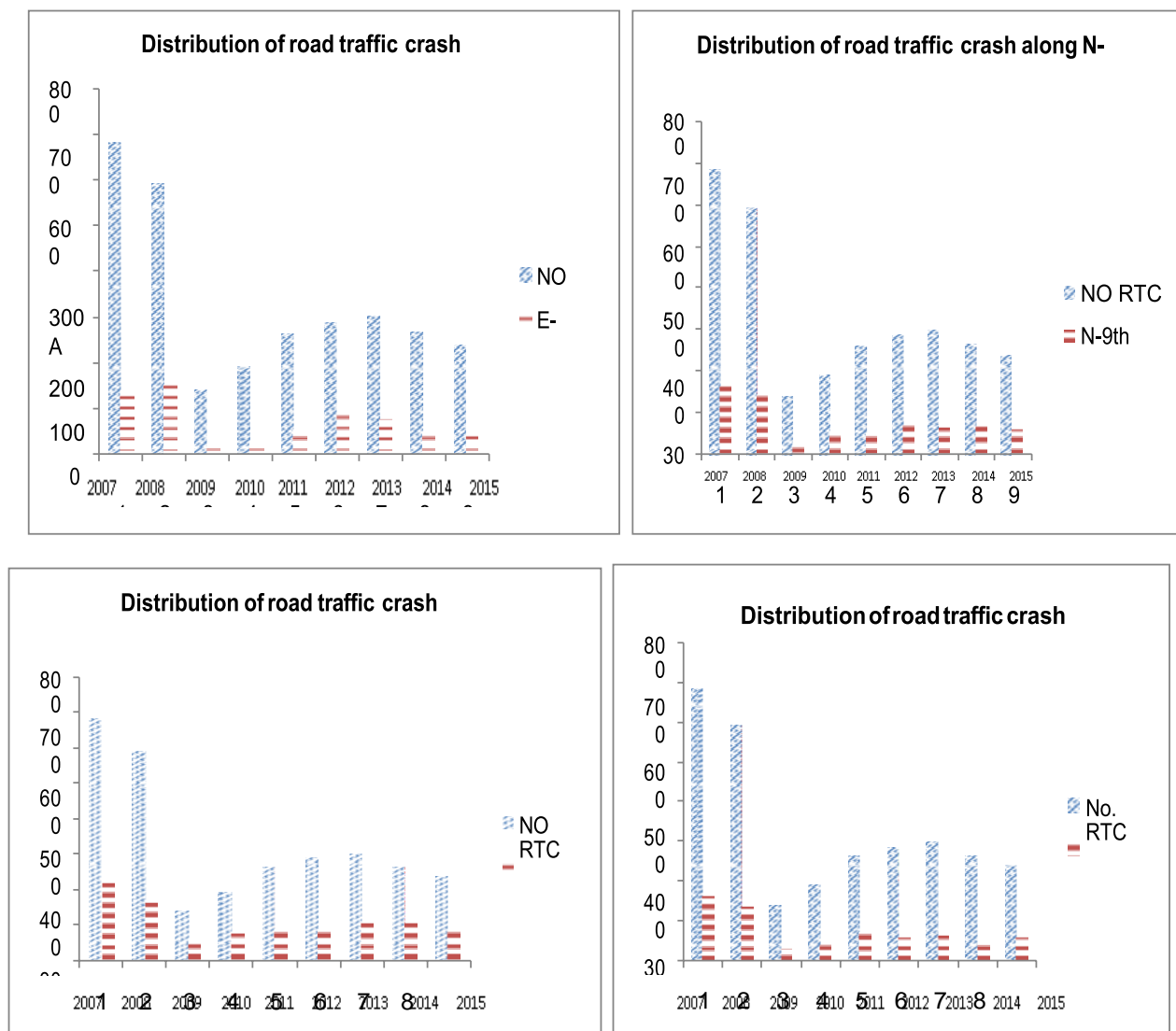
The data used in this research work is the monthly road traffic crashes sourced from the Federal Road Safety Commission Enugu State Nigeria. The data set contains a time series length of 108 starting from year 2007 to year 2015.

Distribution of Road Traffic Crash along the Study Routes

The FRSC record showed that 2,959 Road Traffic crash (RTC) cases were reported with 629,697,983 and 650 crashes along ENU-ABK, NSK-9th Mile, ENU-ONT and ENU-PH roads respectively. The following table shows the distribution of Road Traffic Crash along the Routes.

Table 1: The distribution of road traffic crash along the selected routes

YEAR	NO RTC	ENU-ABK	% NSK-9 th Mi	%	ENU-ONT	%	ENU-PH	%
2007	683	138	20.2 163	23.9	220	32.2	162	23.70
2008	592	151	25.5 142	24.00	164	27.7	135	22.80
2009	139	20	14.4 29	20.9	60	43.2	30	21.58
2010	190	28	14.7 50	26.3	74	38.9	38	20.00
2011	262	52	19.8 53	20.2	89	34.0	68	25.95
2012	288	81	28.1 68	23.6	84	29.2	55	19.10
2013	301	73	24.2 66	21.9	103	34.2	59	19.60
2014	265	45	17.0 68	25.7	104	39.2	48	18.11
2015	239	41	17.2 58	24.3	85	35.4	55	23.10
Total	2,959	629	697		983		650	

Fig 3: Multiple bar charts showing the distributions of road traffic crash along the routes**Table 2: The descriptive statistics of the selected routes**

Variables	Mean	Standard deviation	Variance	Minimum	Maximum	Skewness	Kurtosis
ENU-ABK	5.84	4.973	24.726	0	25	1.542	2.698
NSK-9THMILE	6.45	4.699	22.082	0	27	1.679	3.641
ENU-ONIT	9.10	5.508	30.335	2	30	1.790	3.600
ENU-PHC	5.82	4.271	18.240	0	22	1.66	2.95

Considering the mean number of accident cases in each of the routes, it is obvious that accident occur more in ENUGU-ONITSHA road than the other three routes (ENU-ABK, NSK-9THMILE and ENU-PHC). The number of accident cases in a month within the period under study fluctuates between 0 and 25, 0 and 27, 2 and 30, and 0 and 22 for ENU-ABK, NSK-9THMILE, ENUGU-ONITSHA and ENU-PHC routes respectively. The rate of variability in the monthly number of accident cases in ENUGU-ONITSHA route is explosive as it has the highest variance (30.335). The skewness taking a positive value for each of the routes indicates that more of the number of accident cases in a month cluster to the left of their means (5.84, 6.45, 9.10 and 5.82). On the other hand, the kurtosis values indicate that the distribution of each of the routes is normal (since the kurtosis values are each greater than or equal to 3 by approximation).

3.3 Test for Stationarity

Our interest here is to carry out a test to determine whether the series are really stationary or not. The Augmented Dickey Fuller test (ADF) for unit root was conducted to determine the order of differencing that will make the series stationary. To illustrate this, consider an autoregressive model of order 1 without constant given as

$$z_t = \phi z_{t-1} + e_t,$$

$$e_t \sim N(0, \sigma^2).$$

We try to test the null hypothesis

$H_0 : \phi = 1$ (unit root exist, the series is non-stationary) against the alternative $H_1 : |\phi| < 1$
(Unit root does not exist, the series is stationary)

The null hypothesis has the implication that the series are not stationary but can be made stationary by differencing the series at order 1, while the alternative asserts that the series are already stationary. Using the augmented dickey-fuller test, we consider the test-statistic given as

$$t_{ADF} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

is the model parameter and $SE(\hat{\phi})$ is the standard error of the parameter.

From the test carried out, each of the series was made stationary at the first differencing. The result of the test for unit root is tabulated below:

Table 3: The result of the augmented Dickey-Fuller test

ENU-ABK		
	ORIGINAL SERIES	DIFFERENCED SERIES AT ORDER 1
ADF	-2.7009	-5.7953
P-VALUE	0.2859	0.01
NSK-9THMILE		
ADF	-2.5436	-5.2991
P-VALUE	0.3512	0.01
ENU-ONITSHA		
ADF	-3.0543	-7.1643
P-VALUE	0.1393	0.01
ENU-PHC		
ADF	-2.9827	-6.2256
P-VALUE	0.1690	0.01

From the table 3 above, each of the original series is not stationary. This can be seen from the P- value of the test for the original series which are so big that the null hypothesis cannot be rejected. However, the P-values of the test for stationarity of the series differenced at order 1 are very small, indicating that each of the series at difference of order 1 is stationary.

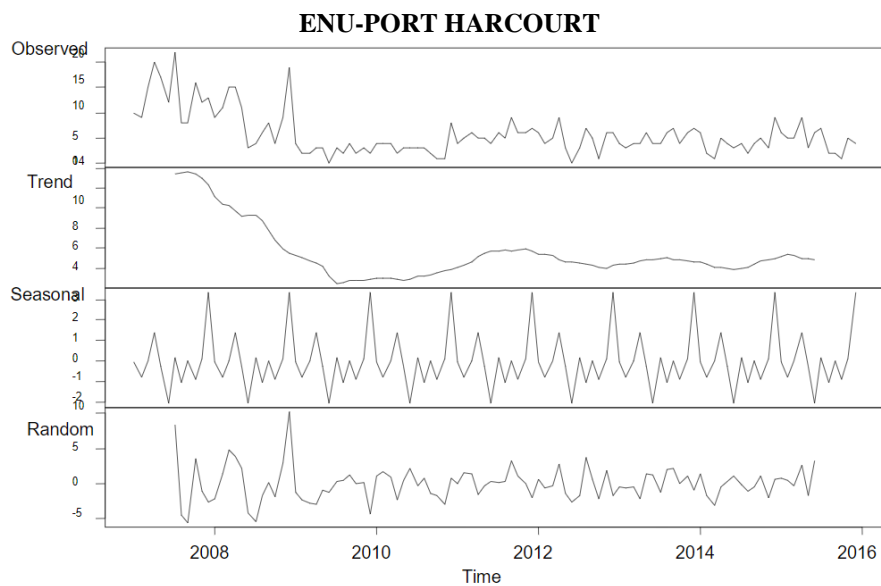
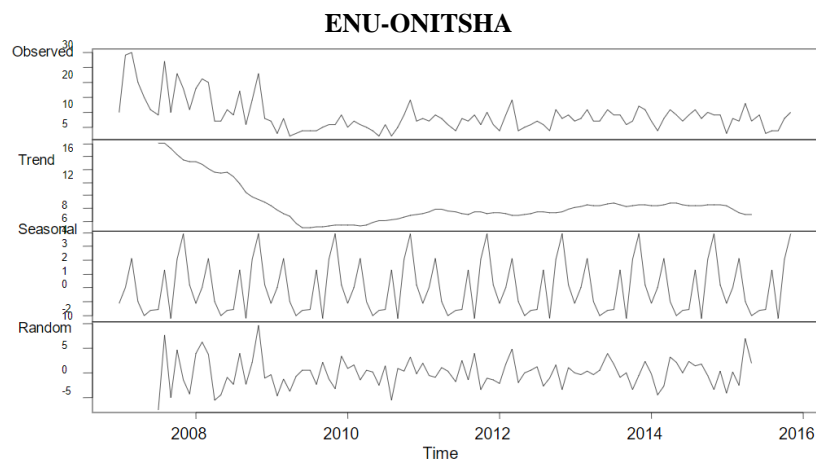
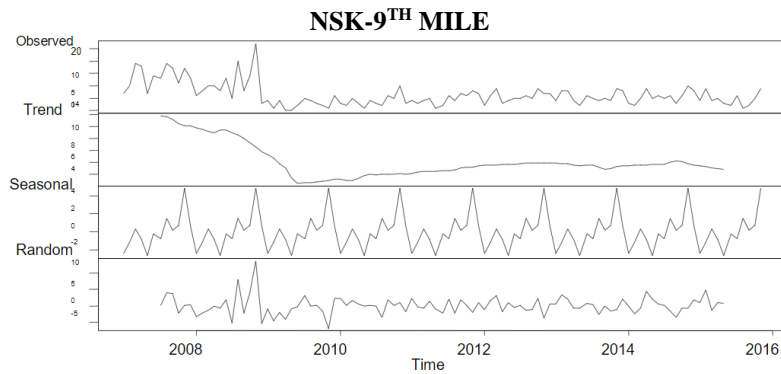
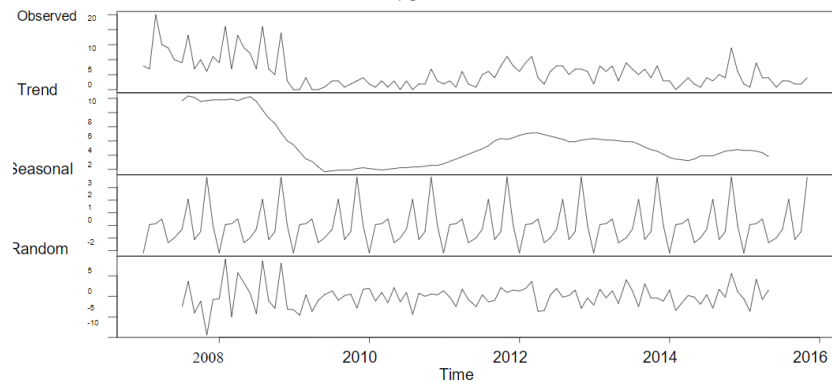
Decomposition of the Observed Series

Now, we want to break down the series into components: trend, seasonal and random effects, using additive model. Prajakta (2014) asserts that in the plots of the series, the distinguishing characteristic between these two types of time series models (additive and multiplicative) is that in the additive case, the series shows steady seasonal fluctuations, regardless of the overall level of the series; where as in the multiplicative case, the size of the seasonal fluctuations vary, depending on the overall level of the series. In addition to this observation, the seasonal components after decomposition of the series take regular pattern in additive case but irregular fluctuation in multiplicative case. Therefore, the regular pattern of the seasonal effects as depicted in the decomposition graph informed the use of additive model, $Z_t = T_t + S_t + E_t$ to describe the series with the seasonality index tabulated below.

Table 5: The computational results of seasonality index along the routes

Period	Jan.	Feb.	March	Apr	May	June	Jul	Aug	Sept.	Oct.	Nov.	Dec.
Enu-Abk	0.139	-1.653	-0.465	0.264	0.306	-1.736	-1.049	0.514	0.951	-0.861	-0.424	4.014
Nsk-9 th mi	1.516	-1.984	-0.984	-0.296	-0.401	-2.922	0.328	-0.526	0.203	0.391	0.557	4.119
Enu-Onit	0.414	-1.043	0.123	1.186	-0.731	-1.543	-1.898	-0.085	0.332	-1.668	2.102	2.811
Enu-Ph	0.359	-0.767	-0.182	0.484	-0.182	-1.245	-0.017	-0.620	0.526	-0.724	0.234	2.130

Figure 8: Decomposition of the observed series using additive model
ENU-ABK



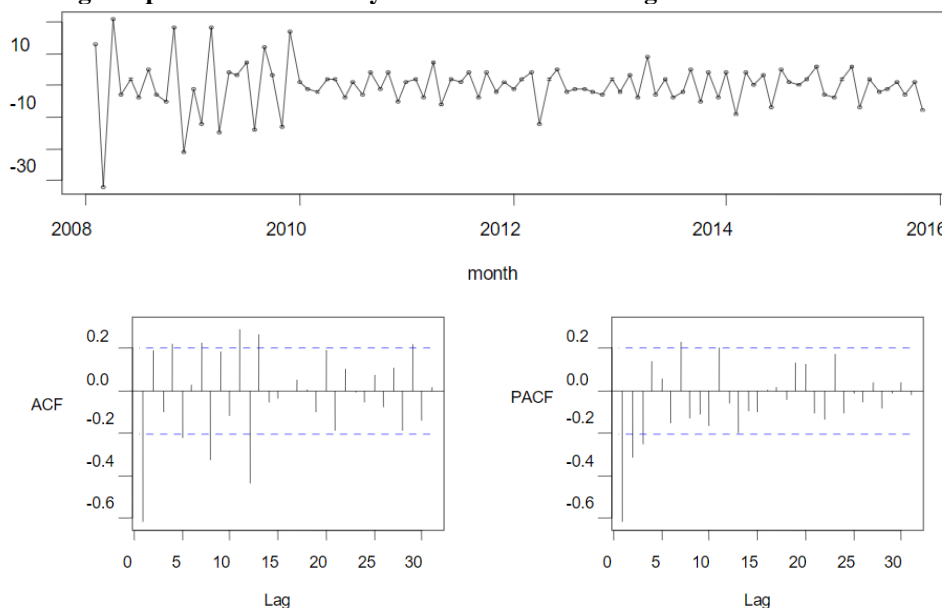
From the decomposition plots, it is clear that each of the series has pronounced seasonal effect. In addition to the information revealed by the plots, it is clearly believed that the rate of road traffic crashes is related in months year to year. It is always very high within festive periods between the month of March and May (Easter period) and November and December (Christmas period), and relatively low in other months over all the years considered. The repeating pattern in the series indicates the presence of seasonal effects and calls for seasonal differencing (that is differencing at order 12) before a SARIMA model is fitted to each of the route. The essence of the seasonal differencing is to retain the seasonal pattern in the long run forecast.

3.5 Seasonal Differencing of Each of the Series

Consider a series with seasonal component S . Before an appropriate SARIMA model can be fitted on the series, it is imperative to do seasonal differencing on the series. That is differencing at seasonal lags 12, 24, 36, etc. With a seasonal time series, it can be made stationary by seasonal differencing which is defined as a difference between an observation and observation with lag that is a multiple of S . For instance, a first order seasonal difference is the difference between an observation and the corresponding observation from the previous year. That is $z'_t = z_t - z_{t-m}$, where z'_t is the seasonally differenced series. For the sake of parsimony, seasonal differencing ought not to exceed that of order two.

FOR ENU-ABK

Figure 9: The correlogram plot of the seasonally differenced series along ENU-ABK route



In the plots of the seasonally differenced series (after the first differencing of the series along ENU- ABK route to make the original series stationary), there is a spike in the ACF at lag 12, but nothing at seasonal lags in the PACF. This is suggestive of a seasonal MA (1) term. In the non-seasonal lags, there are three significant cut-off in the PACF and one significant cut-off in the ACF, suggesting a possible AR (3) term and MA (1) term in the non-seasonal component of the SARIMA model. Consequently, this initial indication suggests that a possible model for these data is an ARIMA (3,1,1)(0,1,1)₁₂. For the sake of parsimony, we refuse to tolerate AR (3) component in the model and consider ARIMA (2,1,1)(0,1,1)₁₂. We therefore fit this model, along with some alterations within the neighborhood, and take note of their information criteria which are shown in the following table:

Table 6: Information criteria for the SARIMA model along ENU-ABK

SARIMA	AIC	AICC	BIC
(2,1,1)(0,1,1)	559.56	560.24	572.33
(1,1,1)(0,1,1)	557.62	558.06	567.83
(1,1,0)(0,1,1)	565.79	566.05	573.45
(2,1,0)(0,1,1)	561.64	562.08	571.85
(1,1,2)(0,1,1)	558.70	559.38	571.47
(2,1,1)(0,1,2)	560.00	560.95	575.32
(1,1,1)(0,1,2)	558.06	558.73	570.82
(1,1,0)(0,1,2)	565.34	565.78	575.55
(2,1,0)(0,1,2)	561.74	562.42	574.51
(1,1,2)(0,1,2)	559.08	560.03	574.40
(2,1,1)(1,1,1)	560.23	561.18	575.55
(1,1,1)(1,1,1)	558.29	558.96	571.06
(1,1,0)(1,1,1)	565.71	566.16	575.93
(2,1,0)(1,1,1)	562.09	562.76	574.86
(1,1,2)(1,1,1)	559.24	560.20	574.56

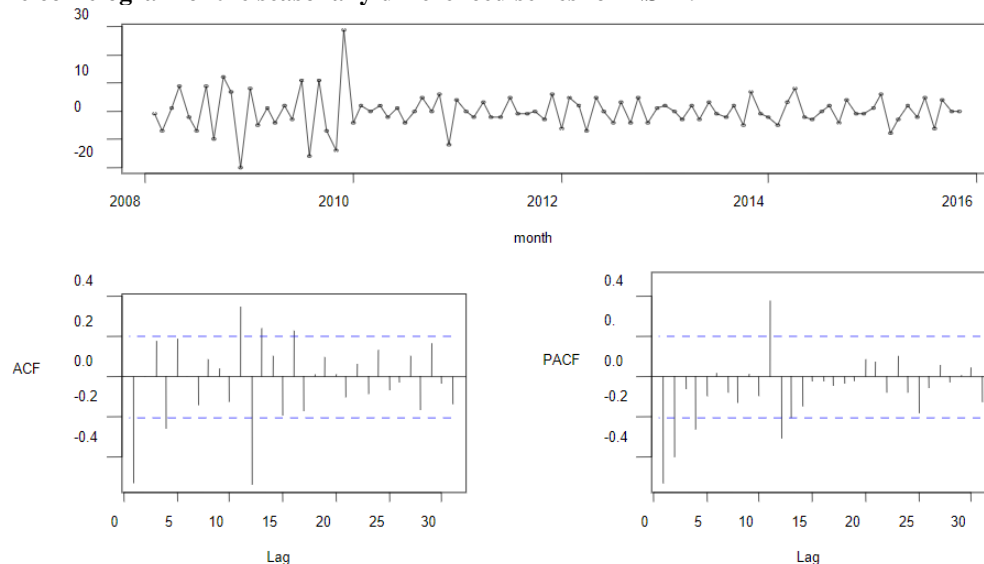
Among all the SARIMA models considered for ENU-ABK route, SARIMA (1,1,1)(0,1,1) was chosen as the best as it has the minimum information criteria considered. The parameters of the model are presented in the table below.

Table 7: The parameter estimate of SARIMA (1,1,1)(0,1,1)

parameters	ϕ_1	θ_1	$\theta_s, s = 1$
Value	-0.2819	-0.5554	-0.9999

FOR NSK-9th MILE

Figure 10: The correlogram of the seasonally differenced series for NSK-9th MILE



Similarly, in the plots of the seasonally differenced data (after the first differencing of the NSK-9TH MILE route to make the original series stationary), there are significant spikes in the ACF and PACF respectively at lag 12 (seasonal lag). These suggest a seasonal AR (1) term and MA (1) term. In the non-seasonal lags, there are three significant cut-off in the PACF and one significant cut-off in the ACF suggesting a possible AR (3) term and MA (1) term in the non-seasonal component. These suggestions gave a possible model for the series as ARIMA (3,1,1)(1,1,1)₁₂. For the sake of parsimony also, we refuse to tolerate AR (3) component in the model and consider ARIMA (2,1,1)(1,1,1)₁₂. We therefore fit this model, along with some alterations within the neighborhood, and take note of their information criteria which are shown in the following table

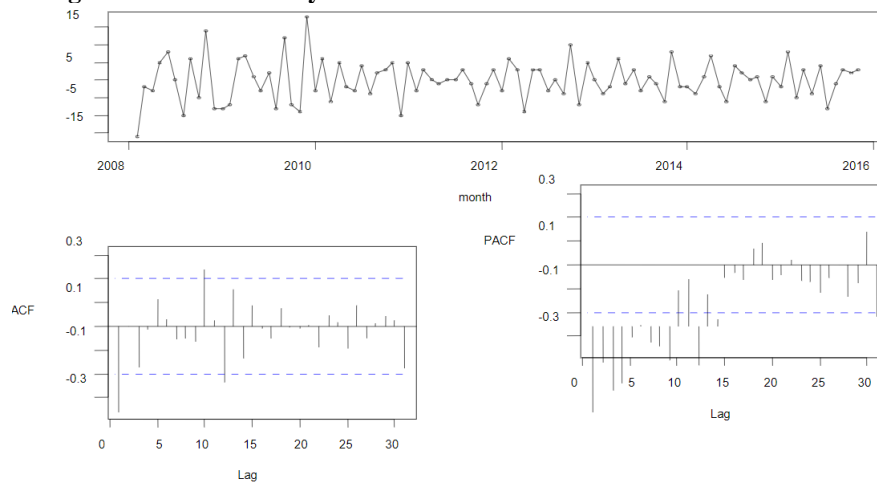
Table 8: Information criteria for the SARIMA model along NSK-9th Mile

SARIMA	AIC	AICC	BIC
(2,1,1)(1,1,1)	536.92	537.59	551.91
(2,1,0)(1,1,1)	537.20	537.87	549.97
(2,1,2)(1,1,1)	538.57	539.86	556.45
(1,1,1)(1,1,1)	534.62	535.29	547.39
(0,1,1)(1,1,1)	533.14	533.59	543.36
(0,1,2)(1,1,1)	534.62	535.29	547.39
(0,1,0)(1,1,1)	576.51	576.78	584.17
(2,1,1)(1,1,0)	536.52	537.19	549.29
(2,1,1)(1,1,2)	538.48	539.77	556.36
(2,1,1)(0,1,1)	537.92	538.59	550.69
(2,1,1)(2,1,1)	538.52	539.81	556.40
(1,1,1)(0,1,1)	535.93	536.37	546.14
(0,1,2)(0,1,2)	535.18	535.85	547.95
(1,1,1)(0,1,2)	535.17	535.85	547.94
(2,1,1)(0,1,0)	575.29	575.74	585.51
(2,1,0)(0,1,0)	578.07	578.33	585.73

From the above table, SARIMA (0,1,1)(1,1,1) is the best model for NSK-9TH MILE route with the following parameters

Table 9: The parameter estimate of SARIMA (0,1,1)(1,1,1)

parameters	θ_1	$\phi_s, s = 1$	$\theta_s, s = 1$
value	-0.7423	-0.3734	-0.3785

FOR ENU-ONT**Figure 11: The correlogram of the seasonally differenced series for ENU-ONT route**

In the plots of the seasonally differenced data (after the first differencing of the ENU-ONITSHA route to make the original series stationary), there is a significant spike in the ACF at lag 12, but nothing at seasonal lags in the PACF. This suggests a seasonal MA (1) term. In the non-seasonal lags, there is one significant cut-off each in the ACF and PACF suggesting a possible AR (1) and MA (1) terms. Consequently, this initial analysis suggests that a possible model for this series (ENU-ONITSHA route) is an ARIMA (1,1,1)(0,1,2)₁₂. We fit this model, along with some variations within the neighborhood, and take note of their information criteria which are shown in the following table:

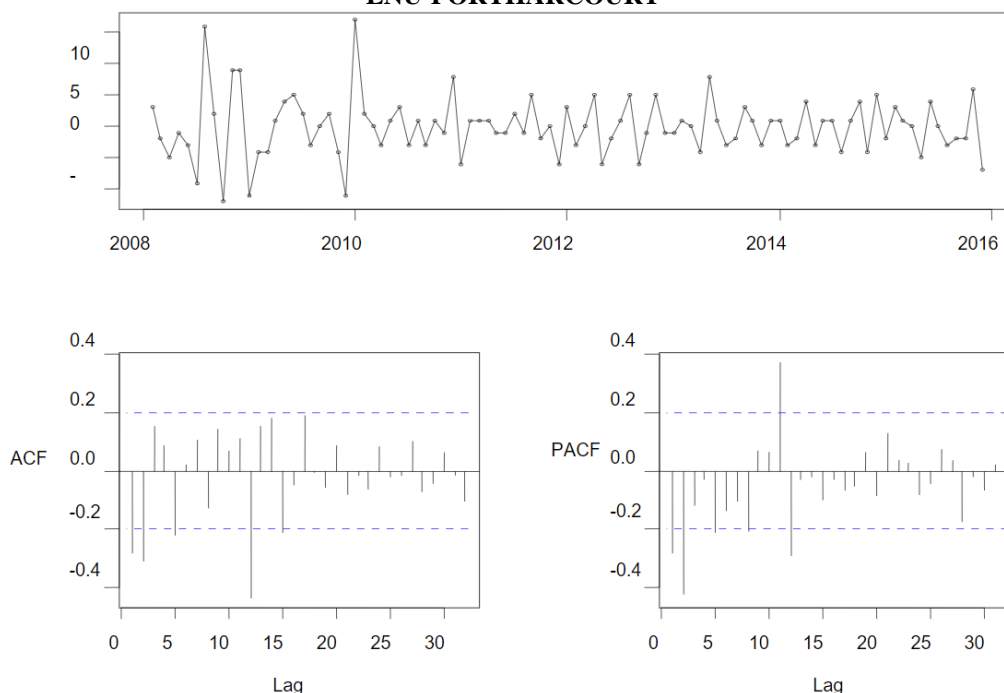
Table 10: Information criteria for the SARIMA model along(ENU-ONITSHA)

SARIMA	AIC	AICC	BIC
(1,1,1)(0,1,2)	558.68	559.35	571.45
(1,1,0)(0,1,2)	577.22	577.66	587.43
(1,1,2)(0,1,2)	560.29	561.24	575.61
(0,1,1)(0,1,2)	556.76	557.20	566.97
(1,1,1)(0,1,2)	558.68	559.35	571.45
(2,1,1)(0,1,2)	560.57	561.52	575.89
(1,1,1)(0,1,1)	556.95	557.40	567.17
(1,1,1)(0,1,2)	558.68	559.35	571.45
(1,1,1)(0,1,0)	566.92	567.18	574.58

The model with minimum information criteria is SARIMA (0,1,1)(0,1,2), therefore, it is taken to be the best model for ENU-ONITSHA route. The parameters of the model are in the table below

Table 11: The parameter estimate of SARIMA (0,1,1)(0,1,2)

parameters	θ_1	$\theta_s, s = 1$	$\theta_s, s = 2$
value	-0.7939	-0.4692	-0.0727

FOR ENU-PH**Figure 12: The correlogram of the seasonally differenced series for ENU-PH route
ENU-PORTHARCOURT**

Also, in the plots of seasonally differenced data (after the first differencing of the ENU-PH route to make the original series stationary), there is a significant spikes in the ACF and PACF. This suggest a possible a seasonal AR (1) and MA (1) term. In the non-seasonal lags, there are two significant cut-off in the ACF and PACF respectively. This suggests a possible AR (2) and MA (2). Consequently, this initial analysis suggests that a possible model for this series (ENU-PH) is an ARIMA (2,1,2)(1,1,1)₁₂. We also fit this model, along with some variations within the neighborhood, and take note of their information criteria which are shown in the following table:

Table 12: Information criteria for the SARIMA model along ENU-PHC

SARIMA	AIC	AIC _c	BIC
(1,1,2)(1,1,1)	504.95	505.90	518.27
(0,1,2)(1,1,1)	507.07	507.74	519.84
(1,1,1)(1,1,1)	508.91	509.59	521.68
(1,1,0)(1,1,1)	524.52	524.96	534.73
(2,1,2)(1,1,1)	506.67	507.96	524.55
(2,1,0)(1,1,1)	510.73	511.40	523.50

The model with minimum information criteria is SARIMA (1, 1, 2)(1,1,1), which is taken to be the best model for ENU-PHC. The parameters of the model are shown below

Table 13: The parameter estimate of SARIMA (1, 1, 2) (1,1,1)

parameters	ϕ_1	θ_1	θ_2	$\phi_s; s = 1$	$\theta_s; s = 1$
value	-0.7409	0.2723	-0.7277	-0.1922	-0.6722

3.6 Diagnostic checking

To confirm the goodness of these chosen models, we examine whether there is serial correlation among the residuals of the fitted models. When error terms from different time periods (or cross section) are correlated, we say that the error term is serially correlated. Our interest is to check whether such serial correlation exist among the residuals. This is done using the Ljung-Box (1978) test for goodness of fit. The null hypothesis is tested against the alternative which posits that there is serial correlation among the residual of the model.

$$Q_n = T(T+2) \sum_{k=1}^n (T-K)^{-1} r_k^2(e) \sim \chi_n^2$$

Test statistic:

Using the test statistic above, the P-value of (Q_n) for each of the fitted models is presented in the table below

Table 16: The Goodness of Fit Test result

ROUTES	d.f	P-value
ENU-ABK	16	0.6543
NSK-9THMILE	16	0.7707
ENU-ONITSHA	16	0.3924
ENU-PHC	16	0.720

The P-value of all the tests for the presence of serial correlation among the residuals of each of the fitted models is relatively high. This does not support the rejection of the null hypothesis. Therefore, we conclude that since there is no serial correlation among the residuals of the fitted models, the models are each good enough to explain the generation process of the cases of accident in each of the routes.

3.7 Model Performance Measures

Here, we shall make use of some error measures to assess the performance of the fitted models. The following measures are considered.

Table 17: The Error Performance measures along the routes

ROUTES	MAE	SSE	MSE	MAPE
ENU-ABK	2.4785	1347.775	12.4794	∞
NSK-9THMILE	2.4849	1282.806	11.8778	∞
ENU-ONITSHA	2.9345	1707.121	15.8067	0.4732
ENU-PH	2.2016	917.905	8.2399	∞
RTC	6.2168	9014.257	83.4653	0.5498

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary

The seasonal autoregressive integrated moving average model is the standard Box-Jenkins model used for series that exhibit seasonality. The series under consideration were identified to have a consistent seasonal pattern, thus we went on to do seasonal differencing to prevent the seasonal pattern from “dying out” in the long term forecast. The SARIMA modeling started with exploratory data analysis which involved exploration of the series parameters and test for stationarity. The correlogram and sequence plot of the observations captured the behavior of road traffic crashes in the selected routes and revealed the presence of seasonal effect, hence the need for seasonal differencing. The seasonality identified in each of the routes is additive since the amplitude is independent of the general level of the series. The order of the SARIMA models was identified through a critical study of the spikes in the correlogram. Being a multiplicative model made up of both seasonal and non-seasonal components, the cut-offs and spikes in the correlogram guided the identification of the order for the seasonal and non-seasonal components respectively. In an attempt to justify the robustness (goodness) of the fitted SARIMA models, Ljung-Box test for goodness of fit was conducted and it was observed that there is no serial correlation among the residuals of the fitted models.

References

- [1].Adebayo, H.O. (2015). GIS based analysis of road traffic accidents in Nigeria (1960 -2012).
- [2].*Ethiopian Journal of Environmental studies and Management*.8(6), 675- 691.
- [3].Agbeboh&Osabuohien(2013). Empirical analysis of road traffic accidents: A case study of Kogi state, North-central Nigeria..
- [4].Akpoghomeh, S.O (1998). Temporal variations in road traffic accidents in Port Harcourt metropolis, *Journal of Transportation.Studies*. 2(1):14-35.
- [5].Atubi A.O. (2015). Modelling deaths from road traffic accidents in Lagos state.*American Journal of Social Science*, 4(5)
- [6].Atubi, A.O. &Onokala, P.C. (2009).Contemporary analysis of variability in road trafficaccidents in Lagos State, Nigeria.*Journal of African Geographical Review*.28,11-41.
- [7].Atubi, A.O. (2012). Epidemiology of Injuries from Road Traffic Accidents in Lagos State, Nigeria.*International Journal of Science and Technology*, 1(2), 56-75.
- [8].Avuglah R. K, Adu-Poku, K. A. & Harris, E (2014).Application of ARIMA Models to Road Traffic Accident Cases in Ghana.*International Journal of Statistics and Applications*, 4(5), 233-239
- [9].Balogun, et al. (2015). On the model selection of road accidents data in Nigeria: A time series approach. *American Journal of Research communication* 3(5),139-177
- [10]. Box & Jenkins (1970).*Time series analysis.Forecasting and control*. Holden day, san Francisco, C.A. New York.
- [11]. Brockwell, P. J. (1986). *Time series.Theory and Methods*.Springer-Verlag, New York, Berlin Heidelberg, London.
- [12]. Fomby, T. B. (2008). Exponential smoothing models in tax: Department of Economics southern Methodist University Dallas.

- [13]. Hilas, C. S, Goudos, S. K & Sahalos, J. N. (2006). Seasonal decomposition and forecasting of Telecommunication data: A comparative case study. *Technological Forecasting and Social Change*, 73(5), 495-509. Iwok, I. A. (2016). Seasonal modeling of road traffic accident. 6(4)
- [14]. Katleho, D. (2016). Comparative study of Holt-Winter (Triple) exponential smoothing and seasonal ARIMA. Forecasting short-term seasonal car sales in South Africa. *Risk Governance and Control Financial Markets and Institutions*, 6 .
- [15]. Khair, et al. (2014). Prediction of Road Traffic Accidents in Jordan using Artificial Neural Network (ANN). *Journal of Traffic and Logistics Engineering* 2, (2)
- [16]. Ljung, G. M. & Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303
- [17]. Olubomehin, O.O (2012). Road Transportation as lifeline of the economy in Western Nigeria, 1920 to 1952. *African Journal of History and culture* 4(3), 37.
- [18]. Olushina & Mumini (2011). Multivariate regression techniques for analyzing autocrash variables in Nigeria. *Journal of Natural sciences Research*, 1(1)
- [19]. Oppong, R.A. (2012). Statistical analysis of road accidents fatality in Ghana using Poisson regression. Kwame Nkrumah University of Science and Technology.
- [20]. Paul, D. J. (2015). Analysis of Road Traffic Accident Hotspots along Zaria-Kaduna expressway, Kaduna State.
- [21]. Peter, W. R (1960). Forecasting sales by exponentially weighted moving average. *Management Science*, 6, 324-342).
- [22]. Prajakta, S.K. (2004). *Time series Forecasting using Holt-Winters Exponential Smoothing*.
- [23]. Kanwal Rekhi School of Information Technology, India.
- [24]. Pratte David (1998). Road to Ruin: Road Traffic Accidents in the developing world.
- [25]. Quigley, C., Sims, R., Hill, J., Tripodi, A., Persia, L., Pietrantonio, H., Kharat, M. 2012. Transport planning guidelines for vulnerable road user safety in emerging economies. *Procedia- Social and Behavioural Sciences*, 48, 3220-3229.
- [26]. Ridwan, A. S., Adebola, F.B. & Adegoke, N.A. (2016). Cases of road traffic accident in Nigeria: A time series approach. *Mediterranean Journal of Social Sciences*, 17(2)
- [27]. Salako A.J., Adeoke, B.O & Akanmu, T.A. (2014). Time series analysis for modeling and detecting seasonality pattern of auto-crash cases recorded at Federal Road Safety Commission, Osun State. *International Journal of Engineering and Advanced Technology Studies*, .2(4). 25
- [28]. Sumaila, A.F. (2013). Road crashes trends and safety management in Nigeria. *Journal of Geography and Regional Planning*, 6(3), 53-62.