Innovations

Characterisation and Prediction of the mortality caused by road Traffic accidents on Cameroon major roads

Awa Terence Achiri, Gerald Azeh and Innocent Ndoh Mbue*
Laboratoire de recherche Energie, Matériaux, Modélisation et Méthode (E3M)
Ecole Nationale Supérieure Polytechnique de Douala (ENSPD), Cameroon
Université de Douala, BP 2701, Douala, Cameroon
Corresponding author **: Innocent Ndoh Mbue**

Abstract:

This study uses a seasonal autoregressive integrated moving average (SARIMA) model to describe and forecast the mortality from traffic-related injuries in Cameroon from 2000 to 2021.The database was produced by triangulating local data with the World Bank Group's website's assessment of the nation's road safety. Using the Augmented Dickey-Fuller test, the series after differencing and log transformation was found to be stationary (p<.05). The Bayesian information criteria of root mean squared error, mean absolute error, and mean absolute percentage error were used to assess the developed models. The goodness-of-fit of the various models was assessed using the autocorrelation function (ACF), partial autocorrelation function (FACF), and LjungeBox test of residuals. The built models were evaluated using the Bayesian information criterion, Root Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error. Using the autocorrelation function (ACF) and partial autocorrelation function (FACF) of residuals and the LjungeBox test, the goodness-of-fit of the various models was compared. Based on minimal diagnostic statistics (RMSE = 1.375, MAE =0.965, MAPE = 37.69), the SARIMA (3, 1, 3)x(0, 1, 2)12 model was determined to be the most frugal model from the aforementioned analyses based on minimal diagnostic statistics (RMSE = 1.375, MAE = 0.965, MAPE = 37.69). A Box-Ljung test revealed that there was no white noise at the 5% level of significance ($x^2{}_{(df=10)}$ =5.148; p = .881). Furthermore, neither the ACF nor the PACF plots showed any peaks outside of the insignificant zone. The expected death rate shows a tendency to increase over time. In addition to improving road safety, especially through the extension of public transit, policymakers urgently need to assure safe, affordable, accessible, and sustainable transport networks for all, with an emphasis on the needs of those in vulnerable situations.

Key-Words: *1.Augmented Dickey-Fuller test, 2.Autocorrelation functions, 3.Box-Ljung test, 4.Road traffic injury, 5.Mortality, 6.SARIMA,*

1. Introduction

Thousands of people are killed or injured on our roads every day of every year. Millions of people are hospitalized for weeks after severe accidents each year, and many will never be able to live, work, or play as they once did because of the rising levels of morbidity and mortality associated with road traffic accidents (RTAs). The cost of RTAs is estimated to be between 1–2% of gross national product in low-and-middle-income countries, which is over $100 billion a year (Bishai et al. 2006). Current trends show that if urgent action is not taken, road traffic injuries could be the seventh leading cause of death by the year 2030, and 90% of these deaths will occur in low and middle-income countries (Jacobs et al. 2000).

By the year 2030, SDG3's Target 6 intends to reduce traffic-related injuries and fatalities by 50 percent (Lepojevi and Pei 2011). Additionally, SDG 11's target 11.2 calls for the development of safe, affordable, accessible, and sustainable transportation systems for all, as well as an increase in road safety, particularly through the expansion of public transportation, with a focus on the needs of those who are most at risk, such as women, children, older people, and people with disabilities. However, attaining this important objective is significantly hampered by the current accident patterns in many low- and middle-income nations.

Accurate forecasting of road traffic injury mortality serves as an early warning system, allowing informed decisions and action to be taken to avert potential catastrophes caused by road accidents. By simulating interrupted time series when implementing new traffic enforcement interventions and regulations in the future, it is possible to predict traffic accidents, injuries, and deaths. Several researchers have developed models to predict mortality caused by traffic injuries. For example, commonly used forecasting methods include Holt-Winters and seasonal regression (Tay et al. 2018), time series models (Jain et al. 2018; Erol et al. 2012), multiple linear regression (Amber et al. 2018), (seasonal) autoregressive integrated moving average (SARIMA/ARIMA) (Wai et al. 2019), artificial neural network (ANN) (Imran and Abraham 2007), and the Poisson Generalized Linear Model (Quddus 2008). Nonlinear relationships in data can be effectively extracted using models based on artificial neural networks. Because of their robustness, fault tolerance, and adaptive learning ability, they have been widely used in time series predictions. Unlike the ARIMA/SARIMA, neural network models have nonlinear functions that constitute the linkage between the value at a time (t) and its previous value at (p) (Xujun et al. 2015).

Notwithstanding, ARIMA models are well-known for their forecasting accuracy and efficiency in representing various types of time series with simplicity, and authors such as Akhtar et al. (2018) have stated that they are the best crash predictive models for aggregated time series count data. This assertion is consistent with the findings of other authors such as Xujun et al. (2015), who developed a SARIMA model for monthly road traffic fatalities in China from 2000 to 2011, Akhtar and Ziyab(2018), who discovered that the SARIMA model best fits the prediction model for monthly road traffic injuries in Kuwait from 2003 to 2009, and Weisent et al. (2010), who showed that SARIMA outsmarts the Poisson regression in terms of accuracy for a given time series data. However, the application of the SARIMA model in longitudinal studies requires voluminous or large monthly data sets, which are often difficult to document (Aguero-Valverde and Jovanis 2006).

While most studies on road traffic injury mortality have been effective in developing countermeasures that have helped to reduce road traffic crashes in developed countries (e.g., Mohan et al. 2020) this has not been the case in most developing countries, particularly African countries south of the Sahara, where road traffic crash statistics leading to mortality are still on the rise. According to the World Bank Group (2022), Cameroon has the worst road safety profile in both the central African sub region and among middle-income countries. The few Cameroonian literature studies on the subject are skewed toward the prevalence and pattern of lower extremity injuries caused by road traffic accidents, trends, and contributing factors (e.g., (Ginyu et al. 2021; Tsala et al. 2021). There is still a scarcity of literature on studies on developing statistical models for forecasting monthly average mortality caused by road traffic injuries in the country to assist decision makers in determining appropriate traffic management systems and even acquiring infrastructure. Such time series models could be useful for analyzing and forecasting future mortality rates. Hence, the purpose of this research is to add to the existing body of knowledge on the subject by first of all characterizing the accidents and injuries, then developing an accurate deterministic model based on seasonal autoregressive integrated moving averages to forecast mortality from road traffic injuries in Cameroon.

The Box-Jenkins (1976) type stochastic process was used because it provides a convenient framework that allows an analyst to find an appropriate statistical model that could be used to answer relevant questions about data, and because it has the capability of dealing with assumptions about system structures in a stringent fashion. Accuracy was assessed using Bayesian information criteria (BIC), mean absolute percentage error (MAPE), root mean square error (RMSE) and the Mean Absolute Error (MAE).

2. Materials and methods

We constructed a seasonal ARIMA model, which can combine seasonal differences with non-seasonal differences, and is suitable for analyzing trends and complex seasonal rules. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model:

$$ARIMA \quad \underbrace{\overset{\displaystyle (p,d,q)}{\overset{\uparrow}{Non-seasonal\ part}}} \; \underbrace{\overset{\displaystyle (P,D,Q)_s}{\overset{\uparrow}{Seasonal\ part}}}$$

Where,

p = non-seasonal AR order,
d = non-seasonal differencing,
q = non-seasonal MA order,
P = seasonal AR order,
D = seasonal differencing,
Q = seasonal MA order, and

2.1. Data Collection

This cross-sectional study uses secondary data on mortality caused by road traffic injury obtained from the global status report on road safety published by the World Health Organization(WHO,2018)triangulated and updated with local data obtained from road safety officials (police and gendarmes) from the year 2000 to 2021. First, a review of relevant contributions from the existing body of the literature to identify the theoretical foundation for the research, the level of novelty and relevance of the research, and to help clarify and refine the focus and research questions addressed in the study.

In the second stage, a total of 264 observations of mortality database were constituted. The data was triangulated with a database registered by government agencies. The latter were built up from daily reports send by the various regional delegation of road safety of the national police/Gendarmerie from the all the ten (10) regions of Cameroon. Each accident reported in the location, victims involved, number injured, number of deaths and the damage caused. Three types of accidents are always distinguished; material, corporal, and mortal accidents. The data exploited as indicated are the overall national statistic on road traffic accidents.

2.2.3. Statistical Data Analysis

The prediction model for the mortality caused by road traffic injury was developed based on Box et al. (2008) methodology for Seasonal Autoregressive Integrated Moving Average (SARIMA) model due to its versatility and well-founded theory. The construction of the ARIMA model used in this research consists of four steps (Figur 1)

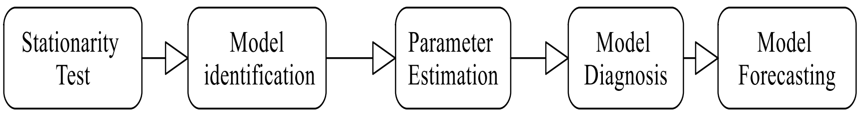


Figure 1. Steps involved in SARIMA model development

A stochastic process, $x_0,\ x_1, \ldots$, is stationary if for any fixed $p \in N, p(x_t, \ldots, x_{t+p})$does not change as a function of t. In particular, statistical properties like mean, variance, and covariances do not vary with time or these stats properties are not the function of time. In other words, stationarity in time series also means series without a trend or seasonal components. One nonseasonal difference (d=1) and one seasonal difference (D=1) were used to stabilize the series (Equation. 1):

$$(1-\beta^{12})(1-\beta)X_t = (X_t - X_{t-1}) - (X_{t-12} - X_{t-13}) \quad (1)$$

Mortality varies in an annual cycle, so s = 12. The model degenerates into an AR model when p is the only nonzero constant and into a moving average (MA) model when q is the only nonzero constant. SARIMA is applied in the present study because mortality exhibits a seasonal pattern.

2.2.4. Model Identification

Using the stationary series' autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, we fixed parameters (p,q) to develop plausible models. A stationary time series has statistical properties such as mean, variance, and autocorrelation that remain constant over time. The ACF is a statistical tool that determines whether earlier incidences in a series are related to later incidences. The PACF measures the amount of correlation between the incidence at time t and the incidence at time t+k after removing the linear dependence from time t+1 to time t+k-1.

Denoted by, ρ(l), the autocorrelation coefficient between a time series, $\{X_t\}$and $\{X_{t-1}\}$ is defined as (Equation 2)

$$\rho(l) = \frac{Cov(X_t, X_{t-1})}{Var(X_t)} \qquad (2)$$

When the ACF and PACF plots are trailing, p= q= 1.
In addition to the ACF and PACF test, the stationarity of the mortality time series data was checked by using the Augmented Dickey–Fuller test, ADF (Cheung and Lai 1995) (Equation 3)

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{i=1}^{k} \varphi_i \Delta Y_{t-i} + \beta t + \epsilon_t \qquad (3)$$

Where

Y_t represents the response variable (mortality rate),
ΔY_{t-i} is the time lagged change in the response variable.
ϵ_t is the white noise error term, t is the time trend.

There is a unit root for the series if p>α, the level of test in the Augmented Dickey–Fuller test. The presence of a unit root shows that the series is non-stationary, and it could be made stationary mostly by applying differencing. Once the stationarity is achieved, the next step is to determine the orders of the autoregressive (AR) and moving average (MA) terms using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

The reason for assuming stationarity is to provide a valid basis for forecasting. A stochastic process is called stationary in the broad sense if it satisfies:

$$E(X_i) = \mu; E[(X_i - \mu)^2] = \delta^2;\ Cov(X_i - X_{i+k}) = y(k) \qquad (4)$$

Where μ & δ^2 are the the population mean, and variance respectively

2.2.4. Parameter estimation

The maximum likelihood approach was used to estimate the parameters of the identified model and the t-values were used to check if the model generated is statistically significant or not. In this study, many ARIMA models were examined and the lowest Bayesian information criterion (BIC), Equation 5, was taken for the optimal model (Box et al. 2015).

$$BIC = n\left[ln\left\{\frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2\right\}\right] + k[\{\ln(n)\}] \qquad (5)$$

Where,

n = The number of data points/observations, or the sample size;
k = The number of estimated input variables,
$\frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})^2$ =$\hat{\sigma}^2{}_e$, the error variance, and

x = The observed data (here, mortality rate)

In addition, the Mean Absolute Percentage Error (MAPE), the Mean Absolute Error (MAE), and the Root Mean Squared Error (RMSE) (Equations. 6, 7, and 8, respectively) were used to assess the models.

Also known as the mean absolute percentage deviation (MAPD), the mean absolute percentage error (MAPE), (Equation 6) is a measure of a forecast system's accuracy.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| * 100 \qquad (6)$$

$$\text{MAPE:}\begin{cases} \text{MAPE} < 5\%; \textit{forecast is reasonably} \\ \text{accurate} \\ 10\% < MAPE < 25\%; \textit{low but acceptable} \\ \text{accuracy} \\ \text{MAPE} > 25\%; \textit{so low, forecast unacceptably} \\ \text{inaccurate} \end{cases}$$

On the other hand, MAE (Equation 7) expresses the precision in the units of the data analyzed.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \qquad (7)$$

A value of this measure close to zero indicates a high precision of the model. Typically, the fitting process is governed by the principle of parsimony, which states that the best model is the one with the fewest parameters that adequately reflects the data.

The Root Mean Square Error (RMSE), (Equation 8) is a standard method for calculating a model's error in predicting quantitative data, and it is regarded as an excellent general-purpose error metric for numerical predictions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2} \qquad (8)$$

Where:

A_t is the observed value,
F_t is the predicted value, and
n is the sample size.

The lower the RMSE, the better a given model can "fit" a dataset.

2.2.5. Model diagnostic checking

The adequacy of the model, considering the properties of the residuals, was checked using the residuals ACF and PACF, and the LjungBox statistics (Q*) (Ljung and and Box. 1978) (Equation 9)

$$Q = T(T+2)\sum_{k=1}^{s}\frac{r^2{}_k}{T-K} \qquad (9)$$

Where

T= number of observations
s = length of coefficients to test autocorrelation
r_k= Autocorrelation coefficient (for lag k)

At least one value of (r) is statistically distinct from zero at the specified significance level if the sample value of Q exceeds the critical value of a $\chi 2$ - distribution with (s) degrees of freedom.

Autocorrelation on the other hand refers to how correlated a time series is with its past values whereas the ACF is the plot used to see the correlation between the points, up to and including the lag unit. In ACF, the correlation coefficient is in the x-axis whereas the number of lags is shown in the y-axis. The Autocorrelation function plot let us know how the given time series is correlated with itself.

2.2.5. Forecasting

Once the final or optimal SARIMA model was found, it was then used to make predictions on the future time points for mortality caused by road traffic injury. Prediction intervals based on the forecasts were also constructed. ARIMA modeling was developed using the Statistical Package for the Social Sciences (SPSS) version 20 software, with all test levels at $\alpha = .05$. To make 12 monthly forecasts, the data were divided into two parts, comprising 264 sample observations from January 2000 to December 2021 and 48 out-of-sample observations from January 2022 to December 2025. The first part is considered as historical period (for fit) and the second part is named the validation period (for forecasting) to verify the out-of-sample accuracy and adequacy of the model for the data.

3. Results and discussion

3.1. Characterisation of the accidents

Accidents occur at varying locations on the Douala –Dschang road at different places every year causing several deaths and injuries, creating accident blackspots (Fig. 2)

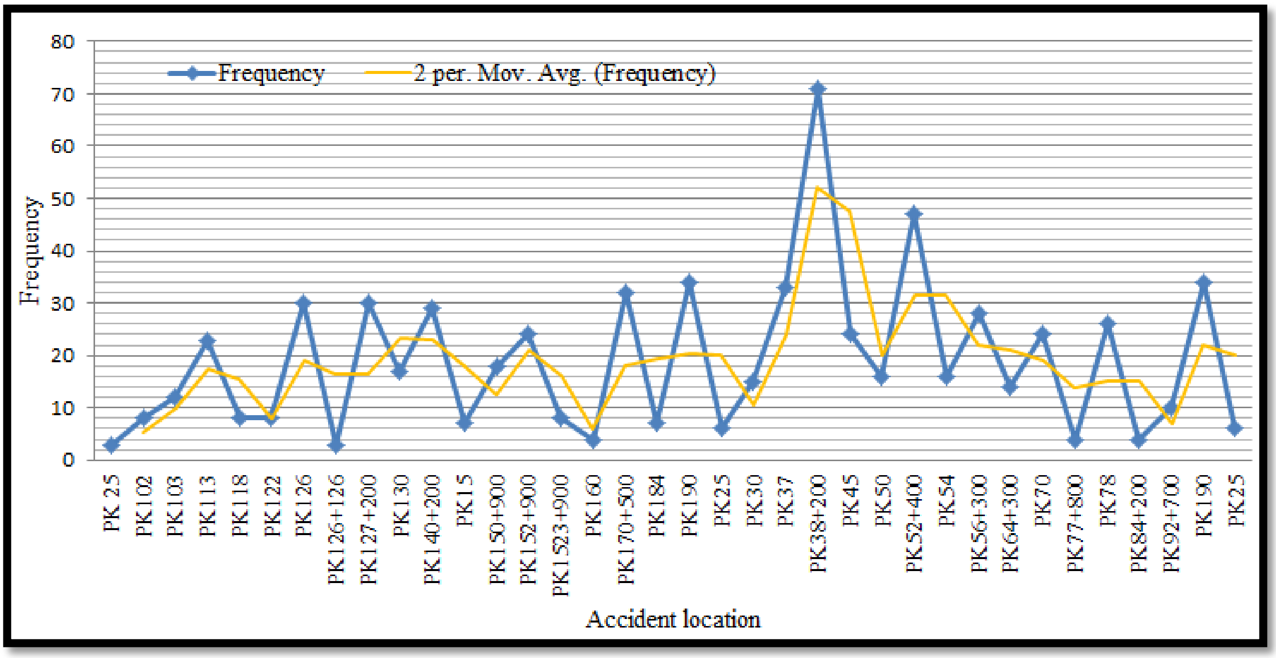


Figure 2: Accident blackspots along the Douala – Dschang road

Most of these accidents are corporal (44%), followed by fatal accidents (30%), and material accidents (26%).Light duty transport vehicles (cars) caused most of the accidents of all types (61.1%) while motor bikes caused the least (5.4%) (Table 1)

Table 1: Classification of accidents according to nature, age category of drivers, seasons, speed limits, and vehicle types

		N	Marginal Percentage
Nature of accident	Fatal	193	30.0%
	Corporal	283	44.0%
	Material	167	26.0%
Age category	Adult	277	43.1%
	Old	110	17.1%
	Teenager	5	0.8%
	Youths	251	39.0%
Season	Dry	447	69.5%
	Rainy	196	30.5%
Speed limit	Excess	314	48.8%
	Required	329	51.2%

Vehicle type	Truck	215	33.4%
	Car	393	61.1%
	Bike	35	5.4%
Total		643	100.0%

Of the 642 data points collected, 148(23.1%) recorded dead, while 494(78.9%) were injuries. The percentage of victims did differ by nature of accident, x^2(2, N = 642) = 374.809, p=0.000, and most of the accidents were more likely to be corporal(Fig 3)

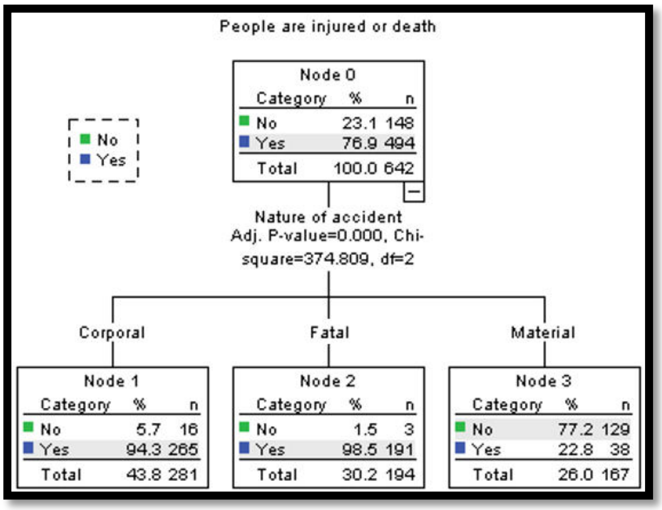


Figure 3: Relationship between injured/dead victims with nature of accident.

Another significant relationship was observed between the nature of accident and type of vehicle involved x^2 (4, N = 643) = 23.114, p =.000. Most light duty vehicles (cars) were more likely to cause accidents than the heavy duty vehicles and motor bikes (Table 2).

Table 2: Relationship between the nature of accident and type of vehicle

			Vehicle type			Total
			Truck	Car	Bike	
Nature of accident	Fatal	Count	73	116	4	193
		% within Vehicle type	34.0%	29.5%	11.4%	30.0%
	Corporal	Count	79	176	28	283
		% within Vehicle type	36.7%	44.8%	80.0%	44.0%
	Material	Count	63	101	3	167
		% within Vehicle type	29.3%	25.7%	8.6%	26.0%
Total		Count	215	393	35	643
		% within Vehicle type	100.0%	100.0%	100.0%	100.0%

Demographically, most of the fatal accidents are caused by young drivers (34.7%), followed by adults (29.2%), the old (22.7%), and the teenagers (0.0%) (Table 3).

Table 3: Relationship between the nature of accident and age category

			Age category				Total
			Adult	Old	Teenager	Youths	
Nature of accident	Fatal	Count	81	25	0	87	193
		% within Age category	29.2%	22.7%	0.0%	34.7%	30.0%
	Corporal	Count	120	42	3	118	283
		% within Age category	43.3%	38.2%	60.0%	47.0%	44.0%
	Material	Count	76	43	2	46	167
		% within Age category	27.4%	39.1%	40.0%	18.3%	26.0%
Total		Count	277	110	5	251	643
		% within Age category	100.0%	100.0%	100.0%	100.0%	100.0%

Youth: <20; Adult: 35-50; Youths: 20-35; Teenagers: <20.

Other explanatory variables were the types of license (x^2 (2, N = 643) = 7.118, p =.028), seasons (x^2 (1, N = 643) = 4.909, p =.027), and speed of the cars (x^2 (1, N = 643) = 6.189, p =.013). Other possible explanatory variables include dilapidated road networks, drunk and driving, over speeding, inefficient enforcement of road safety laws, and most importantly, inadequate to non-existent mortality predictive models, and passenger vehicles.
Several other studies had identified human factors (e.g., health condition of the driver and pedestrian, alcohol consumption, use of mobile phones while driving, distraction due to roadside advertisements, and the age of the driver), to have a more significant influence on the occurrence of road accidents (Shivani and Sebastian, 2019). Caceres et al.2021), like Islam et al.2018) have attributed road accidents to a combination of environmental factors such as fog, ice rain, and rain. Other studies have found that there was a strong relationship between the road accident and the geometrical characteristics of the road, such as sight distance, the radius of curvature, and slope (Mohan et al. 2020). In the same light, Pengaruh et al. (2020) revealed that the road surface condition texture depth and the corresponding skid resistance value have a more significant impact on the occurrence of an accident.

3.2. **Predicting the mortality caused by road accidents**

An exploratory analysis of the mortality in the time series data revealed that the mortality caused by road traffic injuries ranges from 10.2 to 56.16 per 100 000 populations (M = 28.06, SD = 12.69). This recurring problem accounts for 7,810 (4.36% of total deaths) (Bahadorimonfared et al. 2013). The age-adjusted death rate stands at 40.18 per 100,000 people, ranking the country 24th in the world. This correlates with the findings of (WHO 2020), which classified Cameroon's road safety country profile (Figure 4) as the worst both in the central African sub-region and among middle income countries.

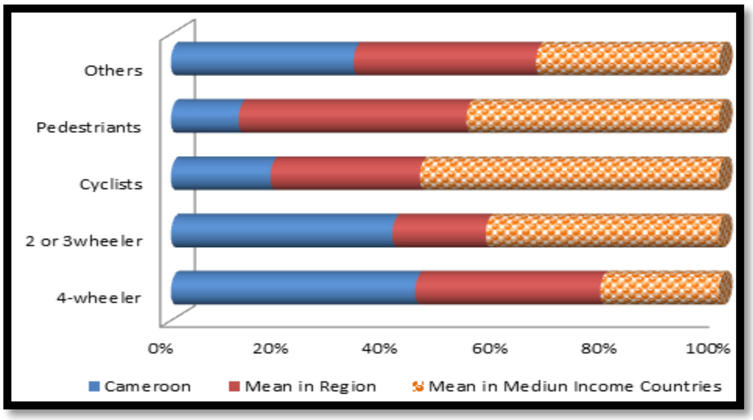


Figure 4: Fatalities by user's comparison chart

The time series data seem to have a deterministic component that is proportionate to the time (Figure 5), suggesting the presence of a time trend.

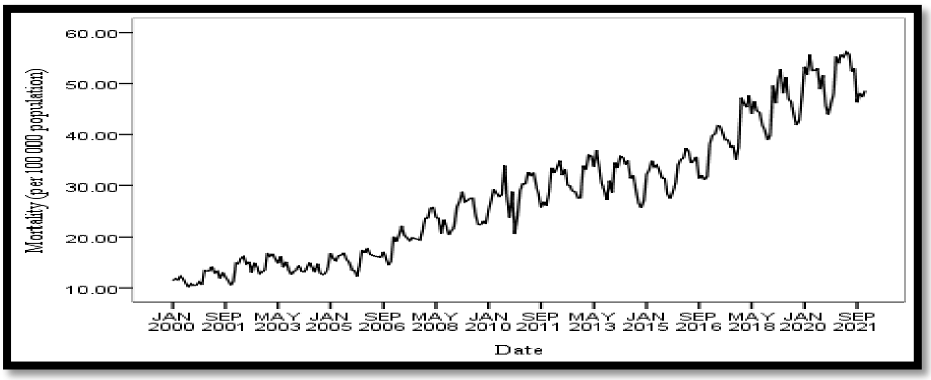
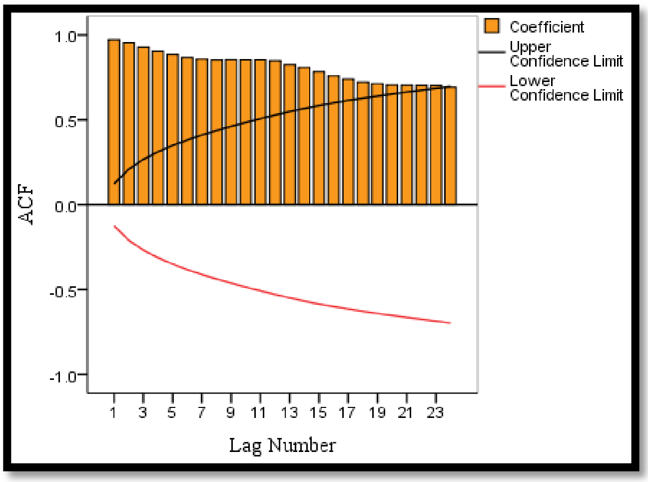
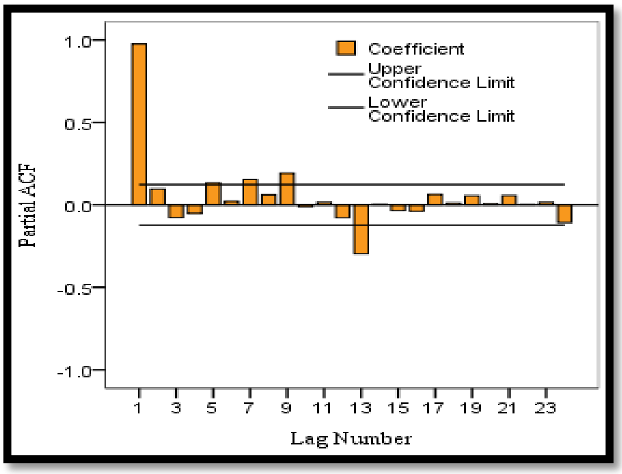


Fig 5: Pattern in mortality caused by road traffic injury

Furthermore, Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) plots of the original series (Fig. 6) failed to die out at high lags -a "scalloped "shape, indicating seasonality (a).





(a) **(b)**

Fig 6: Residual plot of ACF and PACF for mortality, ARIMA (0, 0, 0) with a constant

The PACF plot (Fig 5b) shows significant spikes at lags 1 and 12, and an insignificant spike at lag 24 suggesting seasonality in the data. This was confirmed by the Ljung-Box Q statistic (Q=200.491, DF=18, p-value=.000). Furthermore, the Augmented Dickey-Fuller (ADF) unit root test revealed that the process is not stationary at the 5% significance level, (P = .47>.05) suggesting that he test did not reject the null hypothesis that there is a unit root in the series.

3.2. Stationarity

The Augmented Dickey-Fuller Unit Root Test (ADF = .45>.05). In addition, the "suspension bridge" pattern in the ACF is typical of a series that is both nonstationary and strongly seasonal. Clearly we need at least one order of differencing. Seasonality usually causes the series to be nonstationary because the average values at some particular times within the seasonal span may be different than the average values at other times. For this reason, the variable was log transformed and seasonally differenced (D = 1) of period 12. Differencing ensures that the properties do not depend on the time of observation, eliminating trend and seasonality and stabilizing the mean of the time series. The differenced series (the residuals of a random-walk-with-growth model) looks more-or-less stationary (Figure 7).

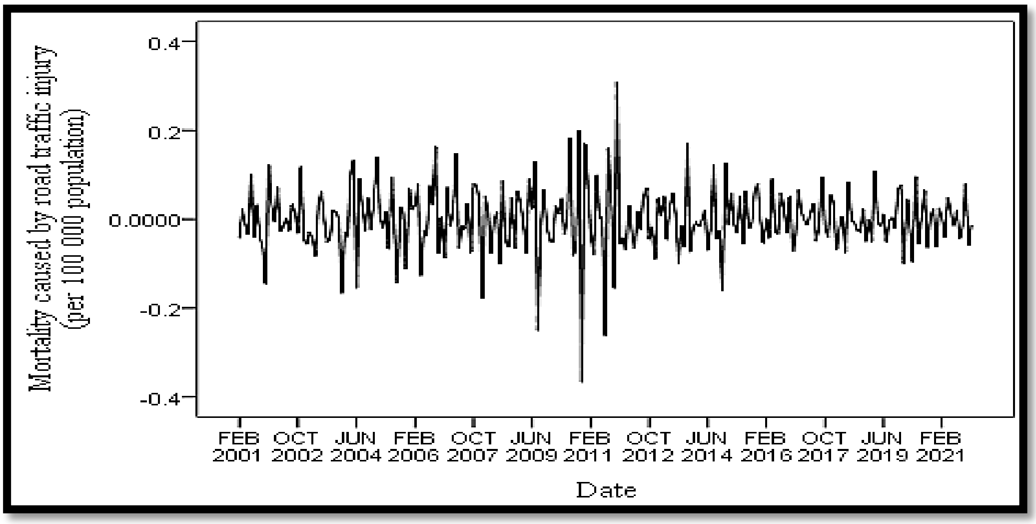
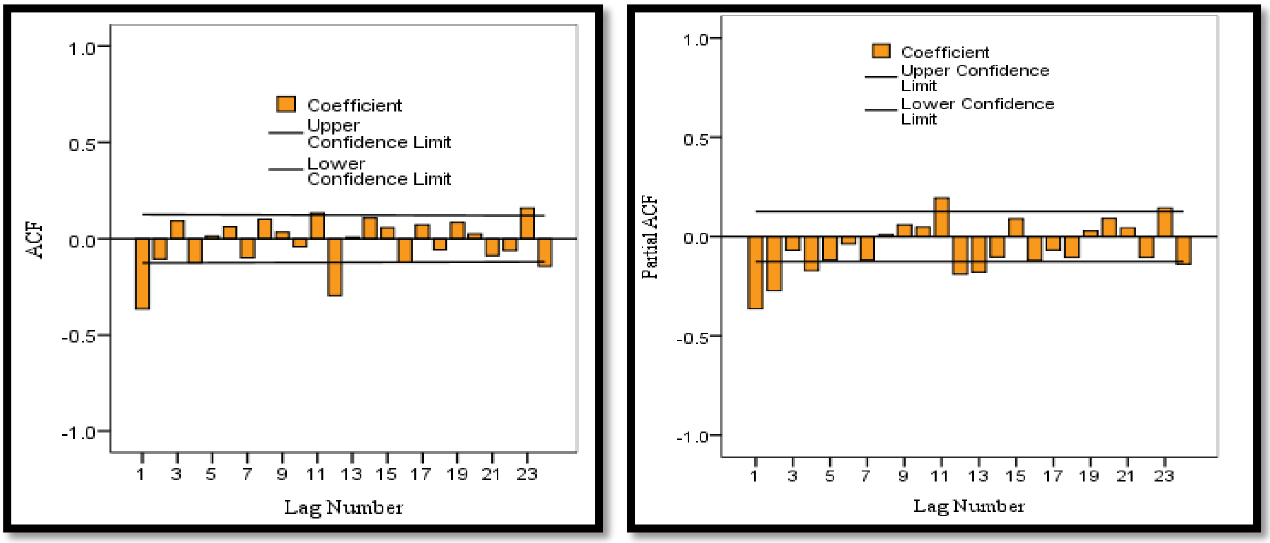


Fig 7: A plot of the natural log-transformed mortality, a difference (1), seasonality (1, period 12)

3.4. Model building

The autocorrelation and partial autocorrelation functions of the log, differenced series (Figure 8) indicated no need for further differencing as they tend to be tailing off rapidly. After nonseasonal and seasonal differencing, Lag 1 now shows a negative autocorrelation, but Lag 2 and the following have mostly insignificant autocorrelations and partial autocorrelations. Lag 12 is due to seasonality.



(a) **(b)**

Fig 8: ACF and PACF plots for the log-transformed seasonally differenced series

We agree with Box and Jenkins (1976) who also suggested that, after differencing, the model might be a moving average of order 1 and a seasonal moving average of order 1. For the first log differenced series, SARIMA (p, 1, q)x(P,1, Q)$_{12}$ are considered where d=D =1 is the order of differencing. The patterns of the ACF and partial autocorrelation function (PACF) plots of the differenced series were examined for the tentative determination of the components of the autoregressive (p) and moving average orders (q) in ARMA(p,q) model. The significant spikes at lags 1 and 2 in the PACF indicate possible non-seasonal AR terms. Therefore, the initial suggestion of an ARIMA (0, 1, 3) x (1, 1, 1)$_{12}$ is proposed. However, studies of neighbouring models to the proposed model suggest an ARIMA (3, 1, 3) (1, 1, 2)$_{12}$, (3, 1, 3) (0, 1, 2)$_{12}$ and, (3, 1, 3) (1, 0, 2) $_{12}$ as the best alternative.

3.5. Model Diagnosis and Selection

The efficiencies of the models to forecast the mortality of caused by road traffic injury was assessed using the Root, Mean Squared Errors (RMSE), Mean Absolute Percentage Error (MAPE). Comparing the various selection criteria indicators, SARIMA (3, 1, 3) x (0, 1, 2)$_{12}$ recorded the least value among all indicators considered (RMSE = 1.375, MAPE = 4.020; MAE = .965). Table 4 presents summary of the results.

Tab 4: Summary of MAE, RMSE MAPE, and Ljung-BoxQ(18) statistic for model estimation

Model	Modelstatistics				Ljung-BoxQ(18)			Outliers
	MAE	RMSE	MAPE	Normalized BIC	Statistics	df	Sig.	
SARIMA(0,1,3)x(1,1,1)$_{12}$	1.046	1.457	40.20	.885	19.044	13	.122	0
SARIMA(3,1,3)x(1,1,1)$_{12}$	.967	1.377	37.74	.838	5.060	10	.887	0
SARIMA(1,1,2) (1,1, 2)$_{12}$	1.053	1.434	4.053	.875	17.818	12	.121	0
SARIMA(3,1,3)x(0,1,2)$_{12}$	.965	1.375	37.69	.836	5.148	10	.881	0
SARIMA(3,1,3)x(1,0,2)$_{12}$	1.020	1.403	40.87	.889	10.237	9	.332	0

Furthermore, a Box-Ljung test reported similar statistic for both model($X^2_{(df=10)}$ =5.14848; p- =0.881) for SARIMA(3,1,3)x(0,1,2)12, and $X^2_{(df=10)}$ = 5.60); p=0.887 for SARIMA (3,1,3)x(0,1,2)$_{12}$ suggesting that there is no evidence of white noises at 5% significance limit for both outperforming models. Furthermore, there were no spikes outside the insignificant zone for both ACF and PACF plots. Therefore, either model is adequate and provides nearly the same three-step-ahead forecasts. However, the study aimed to develop a model that is parsimonious as possible, whilst passing the diagnostic checks. Following this, the SARIMA (3, 1,3) x (0, 1, 2)$_{12}$ model was selected to forecast future readings. A further look at the plots of the residual, ACF, and PACF plot (Figure 9) for the SARIMA (3, 1,3) x (0, 1, 2)$_{12}$ model reveals a random variation- from the origin (0), the points below and above are all uneven, hence the model fitted is adequate.

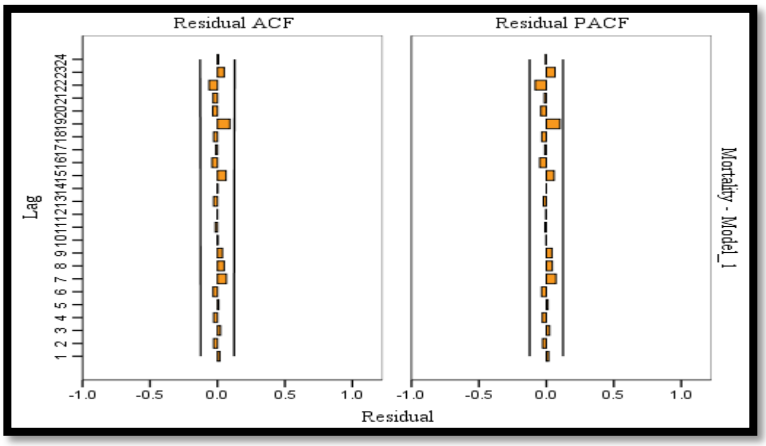


Fig 9: ACF & PACF of the Residuals SARIMA (3, 1, 3) x (0, 1, 2)$_{12}$

Reading from the bottom up, both figures show no pattern in the correlations reported among the residuals nor do any of the correlations extend beyond the vertical 95% confidence intervals included in the plots. This, combined with the Ljung-Box Q statistic, and the Normalized Bayesian Information Criterion (BIC), suggests that the SARIMA (3, 1, 3) x (0, 1, 2)12 model appropriately modeled the dynamics for this time series.
The forecasted figures from SARIMA (3, 1, 3) × (0, 1, 2)12 tend to be very close to the actual data used as test data (Figure 10).

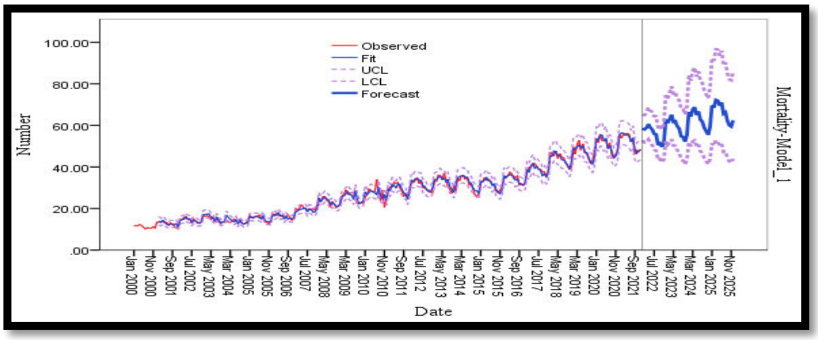


Fig 10: The plot of the SARIMA (3,1,3)×(0,1,2)$_{12}$ forecasted and observed values with 95% CI

The actual figures lay within the 95% confidence interval in most cases, and over 85% of them lay within the forecasted interval. Mortality caused by road traffic injury values (in 100 000 population) are shown by the thicker red sinusoidal curve, the forecasted values are shown by the thick blue line, whilst the bounded light pink shaded region areas show 80% and 95% prediction intervals respectively. The model is validated since the predicted quantity fluctuates around the fit.

The increase in mortality caused by road traffic injury is a major policy concern in international regimes. The current study focuses on the nation of Cameroon which has one of the highest rates of mortality caused by traffic injury in the world. Key challenges that affect road safety and performance strongly correlate with the (Lepojević and Pešić 2011).) suggestions including very slow or lack of improvement of road safety in the country; weak regulatory frameworks and underfunded road safety management at national and local levels; insufficient to total lack of maintenance of road infrastructure with clear road signs and markings; lack of periodic vehicle maintenance; overrepresented motorcycle, which have not been adequately addressed; etc.

SARIMA model was applied to monthly reported mortality caused by road traffic injury data in the country from January 2000 to December 2021 to determine patterns and predict mortality caused by road traffic accident cases in the country. After identifying various tentative models, an efficient model for the mortality cases was selected using appropriate statistics. The adequacies of the model were tested by analyzing standard residuals in different forms. Forty-eight (48) months of forecasts were provided for injury cases. Our results were similar to those obtained (Quddus 2012) who identified SARIMA (1, 1, 2) $(1, 1, 2)_{12}$ model for forecasting road traffic fatalities in Malaysia. However, the model was not as efficient as SARIMA (3, 1, 3) x $(0, 1, 2)_{12}$ that we identified in this study. Elsewhere, (Bahadorimonfared et al. 2013) proved that the SARIMA (0, 1, 1)$(0,1,1)_{12}$ model is appropriate in presenting the seasonal trend of the monthly number of road traffic fatalities in Iran. Similarly, (Mohan et al. 2020) identified SARIMA (1, 1, 1) x $(0,1,1)_{12}$ as efficient models to predict fatalities in China from 2000 to 2011in China. The findings were also consistent with the findings of (Islam et al. 2019), who discovered the SARIMA model to be the best fitting prediction model for monthly road traffic injuries in Kuwait from 2003 to 2009. In contrast, these models failed when applied to mortality caused by car injury in Cameroon. These disparities in the models could be the result of the use of different approaches and different incidences. Data reliability and cleaning procedures could also influence the results of such modeling.

According to the United Nations (UN.2017), if road safety is achieved and related SDG targets 3.64 and 11.25 are met the potential benefits for people will extend beyond their safety. However, road safety cannot be achieved in isolation from other issues addressed in the 2030 Agenda. Countries that have achieved a high level of safety have had to address a wide range of other issues. The Sustainable Development Goals (SDGs, a "road system that is safe, efficient, and meets the transportation needs of all people, for example, facilitates equitable access to education (SDG targets 4.2 and 4.3), health care (target 3.8), and food (target 2.1) and (target 9.1)." A system like this also connects all parts of a country, helping to build economic, social, and environmental ties between cities, peri-urban areas, and rural areas (target 11 a). Countries that have attained a high level of safety have had to deal with a variety of other issues. Countries that have achieved a high level of security have had to deal with a variety of other issues. The same is true for sustainable cities (SDG 11), climate action (SDG 13), and gender issues (SDG 5), all of which should be taken into account when planning transportation for sustainable and equitable solutions. All of these factors contribute to more efficient and long-term improvements in road safety.

4. **Conclusions**

A forecast model was developed in this study to determine the mortality caused by traffic injuries in Cameroon. The forecast model was created using SARIMA, with a mortality rate as the explanatory variable. It was also demonstrated that in an operational scenario, the SARIMA (3, 1, 3)$(0,1,2)_{12}$ model was the best fit for this time series data. It is also important to note that when using the models proposed for forecasting purposes, a transformation in the explanatory variable is required. The conclusions drawn from the study area revealed that, unless otherwise stated, the mortality rate caused by traffic injury in the country generally increased gradually over time. Although longer-term predictions should be treated with caution, the model estimated in this article may provide a reliable approximation of the pattern of growth of the main dimensions of road traffic injury mortality in the country. As a result, these estimations can assist policymakers in monitoring and managing the increased toll that this phenomenon is putting on national public health systems. However, our study suggested that nonlinear relationships may exist among the monthly incidences of mortality so that the SARIMA model did not efficiently extract the full relationship hidden in the historical data. We therefore recommend, sophisticated forecasting techniques such as those based on neural networks could be more useful in future.

References

1. *Aguero-Valverde J., Jovanis P. P. (2006) Spatial analysis of fatal and injury crashes in Pennsylvania Accid. Anal. Prev. 38, (2006), pp. 618–625.*
2. *Akhtar S. and Ziyab A. H. (2018) "Impact of the Penalty Points System on Severe Road Traffic Injuries in Kuwait," Traffic Injury Prevention, vol. 14, no. 7, (2013), pp. 743–748,*
3. *Amber K. P., Ahmad R., Aslam M., Kousar W. A., Usman M., and Khan M. S. (2018). "Intelligent techniques for forecasting electricity consumption of buildings," Energy, vol. 157, pp. 886–893,*
4. *Bishai D., Quresh A., James P., and Ghaffar A. (2006) National road casualties and economic development Health Econ. 15, pp. 65–81.*
5. *Bahadorimonfared A., H. Soori, Y. Mehrabi, A. Delpisheh, A. Esmaili, M. Salehi, and M. Bakhtiyari (2013) "Trends of fatal road traffic injuries in Iran (2004–2011)," PLoS One, Vol. 8, No. 5, e65198,*
6. *Box G.EP, Gwilym M. Jenkins , Gregory C. Reinsel , Greta M. Ljung(2015). Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics, 5th edition, Wiley, 118674928, 9781118674925, 712 pages.*
7. *Box, G., Jenkins, and Reinsel, G. C. (2008), "Time series analysis: forecasting and control," vol. 4, Fourth Edition*
8. *Box, G. E. P., G. M. Jenkins. (1976). Time Series Analysis: Forecasting and control. Rev. ed. San Francisco: Holden-Day*
9. *Cheung Y. W. and Lai K. (1995) "Lag order and critical values of the augmented Dickey-Fuller test," Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 277–280.*
10. *Caceres L., Fernandez M. A., Gordaliza A., and Molinero A. (2021). ´ "Detection of geometric risk factors affecting head-on collisions through multiple logistic regression: improving two-way rural road design via 2+1 road adaptation," International Journal of Environmental Research and Public Health, vol. 18, no. 12, p. 6598.*
11. *Erol A. H. zçelikkan N., Tokgöz A., ze S. L., Zaim S., and Demirel A. F. (2012) "Forecasting Electricity Consumption of Turkey Using Time Series Methods and Neural Networks," 118-127*
12. *Ginyu I. K., Siysi V. V., Asangbeng T. E., Nguedia A., Clement J., and Shey N. D., "Trends and Contributing Factors to Road Traffic Accidents in Douala," American Journal of Epidemiology and Infectious Disease, Vol. 9, pp. 24-31,*
13. *Imran M. and Abraham A. (2007), "Weather analysis using an ensemble of connectionist learning paradigms," Applied Soft Computing 7, pp. 995–1004,*
14. *Islam M. H., Teik Hua L., Hamid H., and Azarkerdar A. (2019). "Relationship of accident rates and road geometric design," IOP Conference Series: Earth and Environmental Science, vol. 357, no. 1, pp. 012040–12110.*
15. *Jacobs, G.D.; Aaron-Thomas, A.; and Astrop, A. (2000). Estimating global road facilities Crow Thorne, U.K. Transport Research Laboratory Report 445, pp. 1-35.*
16. *Jain P. K., Quamer W., and Pamula R. (2018). "Electricity Consumption Forecasting Using Time Series Analysis," in Advances in Computing and Data Sciences, Communications in Computer and Information Science, ch. 33, pp. 327–335*
17. *Lepojevi V. and Pei M. A. (2011) "Forecasting Electricity Consumption By Using Holt-Winters and Seasonal Regression Models," Series: Economics and Organization, vol. 8, no. 4, pp. 421–431*
18. *Ljung, G.M. and Box, G.E.P. (1978) On a Measure of Lack of Fit in Time Series Models. Biometrika, 65, 297-303. https://doi.org/10.1093/biomet/65.2.297.*
19. *Mohan D., Tiwari G., and Bhalla K. (2020) Road Safety in India: Status Report 2020, pp. 1–67, Transportation Research & Injury Prevention Programme, Indian Institute of Technology, Delhi.*
20. *Pengaruh P. M. A., Pmdn T. K., and Dan I. (2020). "Road safety annual report," International Transport Forum, vol. 2507, pp. 1–9.*
21. *Quddus, M. A. (2008). Time series count data models: an empirical application to traffic accidents. Accid. Anal. Prev. 40, pp. 1732–1741. Wai A. H. C., Yi S. S., and Wan F. J. L. (2019)., "Fatality Involving Road Accidents in Malaysia," presented at the Proceedings of the 2019 2nd International Conference on Mathematics and Statistics*
22. *Shivani S. and Sebastian S. (2019). "IoT based car accident detection and notification algorithm for general road accidents, In ternational Journal of Electrical and Computer Engineering (IJECE)," International Journal of Electrical and Computer Engineering, vol. 9, no. 5, p. 4020.*

23. *Tay K. G.. Choy Y. Y., and Chew C. C.(2018)."Forecasting Electricity Consumption Using Fuzzy Time Series," International Journal of Engineering and Technology, vol. 7, 4.30, pp. 342-346..*
24. *Tsala A. S. Z., Makomra V., and Ohandja L. M. A (2021)."An In-Depth Analysis of the Causes of Road Accidents in Developing Countries: Case Study of Douala-Dschang Highway in Cameroon," Journal of Transportation Technologies, vol. 11, pp. 455-470,*
25. *Wai A. H. C., Yi S. S., and Wan F. J. L. (2019)., "Fatality Involving Road Accidents in Malaysia," presented at the Proceedings of the 2019 2nd International Conference on Mathematics and Statistics - ICoMS'19*
26. *Weisent J., Seaver W., Odoi A., and Rohrbach B (2010). "Comparison of three time-series models for predicting campylobacteriosis risk," Infect, vol. 138, pp. 898–906*
27. *Xujun Z., Pang Y., Cui M., .Stallones L. and Xiang H. (2015). "Forecasting mortality of road traffic injuries in China using a seasonal autoregressive integrated moving average model," Annals of Epidemiology, vol. 25, pp. 101-106,.*

Corresponding author's e-mail**: **dndoh2009@gmail.com