A Time-Series Analysis of Traffic Crashes in New York City

Khaled Shaaban
Department of Engineering
Utah Valley University
Orem, UT, USA
kshaaban@uvu.edu

Mohamed Ibrahim
Department of Civil Engineering
Qatar University
Doha, Qatar
mohamed.amin@qu.edu.qa

2022 Intermountain Engineering, Technology and Computing (IETC) | 978-1-6654-8653-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IETC54973.2022.9796728

Abstract— **In New York City, traffic crashes are one of the main causes of fatalities in the city. This study presents a comprehensive time series analysis of road crashes in the city from 2013 to 2019. The crash data were collected, organized, and analyzed at different time levels: yearly, seasonally, monthly and hourly bases. Forecasting of the total number of crashes in the years 2020 to 2025 was conducted using the Box-Jenkins method based on the autoregressive integrated moving average (ARIMA) model. The model was statistically validated using a modified Box-Pierce (Ljung-Box) Chi-Square test. The proposed model was also used for backward prediction of the year 2019 to compare with actual observations. The predicted results showed a good agreement with the actual observed results. The results also showed a strong potential of having a reduction in the total number of crashes in the future.**

Keywords— Fatalities, injuries, ARIMA, driver behavior, accidents

I. INTRODUCTION

Approximately 1.2 million people die every year in traffic-related crashes [1]. This is a problem in developed and developing countries alike [2, 3]. Traffic-related crashes cause great pain and suffering for many families. In addition, the effect of these crashes has resulted in high medical expenses and productivity loss in many societies. While many successful programs resulted in a clear reduction in crash rates, the rates are high, and still considered unacceptable. The problem of traffic crashes is severe due to the complexity of traffic flow and the mixed nature of users (drivers, pedestrians, cyclists, etc.).

In New York City (NYC), approximately 500 people are killed every year due to traffic crashes [4]. From 2012 to 2014, there was an annual average of 1,098 deaths, 12,093 hospitalizations, and 136,913 emergency department visits because of traffic injuries in NYC. Although crashes may not be completely eliminated, proper engineering and management analysis and strategies can reduce crash rates to a certain extent. Therefore, proper investigation and forecasting of crashes can help to recommend countermeasures in terms of design and control.

One of the ongoing efforts to improve traffic safety is research dealing with using statistical analysis to forecast traffic crashes, injuries, and fatalities in the future. Investigators have constantly identified methods to achieve a better understanding of the factors that affect crashes and methods that help to better predict the likelihood of crashes to develop policies and countermeasures to reduce fatalities, injuries, and the number of crashes [5]. Forecasting traffic crashes has historically been utilized as a base for developing traffic safety policies that can help reduce fatalities and injuries. Although the quality of data has not improved as anticipated, the advance in statistical methodologies has permitted researchers to produce better results from existing data [6].

Different techniques have been used to conduct safety analysis. Some of these techniques depend mainly on crash data and others do not [7-10]. When conducting statistical analysis of traffic crashes, different regression techniques were utilized to develop models that can be used for crash prediction. Early models focused on ordinary or normal linear regression. These models are based on assuming a normal error structure for the response variable, a constant variance for the residuals, and the presence of a linear relationship between the response and explanatory variables [11-14].

Some studies used time series to forecast the total number of crashes using autoregressive integrated moving average (ARIMA) models. The results of these studies indicated that ARIMA was a robust model that can be used for forecasting traffic crashes [15-18]. One of the most used time series analysis models in traffic safety is the ARIMA model proposed by Box and Jenkins [19]. Several authors considered the use of the ARIMA models as one of the effective methods of forecasting [20, 21]. The purpose of this study is to conduct a comprehensive analysis of road crashes in NYC using ARIMA models.

II. METHODS

A. Data Collection

Crash data were obtained from the New York City database. The crash data included the information for all crashes that occurred in NYC from 2013 to 2019 with a total of 1.53 million crashes. It should be noted that police reports are required to be filled out for collisions where someone is injured or killed, or where there is at least $1,000 worth of damage

The crash data includes time information for each crash, such as the crash date and the exact crash time. This data has been organized and categorized based on different time levels, namely, the total number of crashes per year, per season, per month, and per hour in the study time duration from 2013 to 2019. It can be noticed that the year 2019 shows an 8.7% decrease in the total number of crashes compared to that in 2018. The data showed that June is the month with the highest number of crashes in the majority of the last six years

B. Time Series

The time series were used to report the total number of crashes that occurred in NYC from 2013 to 2019. Different time levels are used to study the variance of the total number of crashes. Level 1 includes the total number of crashes for each year of the study period (2013, 2014, 2015, 2015, 2017, 2018, and 2019). Level 2 consists of the four seasons (winter, spring, summer, and fall). The winter season includes all the crashes that occurred in the period of (December to February) every year, while the spring, summer, and fall seasons include the crashes in the periods (March to May), (June to August), and (September to November) every year, respectively. Level 3 contains the total number of crashes reported every month (January, February,…., December) for each year of the study period (2013 to 2019).

C. Time Series Forecasting

There are several models to forecast the time series. The best model to forecast future data from a time series is the model that can properly describe the available data. A model which describes the probability structure of a sequence of observations is called a "stochastic process". A time series of N successive observations $z = (z_1, z_2, \ldots, z_n)$ is regarded as a sample realization. This is can be simply using the crash reporting process. A key class of stochastic models is the stationary status. These are assumed to be in a specific form of statistical equilibrium. Particular stationary stochastic status of value in modeling time series are the autoregressive (AR), moving average (MA), and ARMA. Also, another key class of the stochastic processes is the non-stationary status ARIMA models [14]. If the variance of the data occurs in seasonal form, seasonal ARIMA can be used for forecasting. In this study, seasonal ARIMA will be utilized to forecast traffic crashes per month for the years 2020 to 2025. The detailed procedure and discussion of the selected procedure is presented in section

III. ANALYSIS

This section presents the results of the time series analysis and forecasting with related discussion.

A. Time Series Forecasting

Box and Jenkins (1976) have generalized the ARIMA (p, d, q) model to deal with seasonality and defined a general multiplicative seasonal autoregressive integrated moving average (SARIMA) model in the form shown below in Eq (1) and Eq (2):

$$\phi_p(\beta)\Phi_P(\beta^s)W_t = \theta_q(\beta)\Theta_Q(\beta^s)a_t \quad (1)$$

Or:

$$\left(1-\sum_{j=1}^{p}\phi_j\beta^j\right)\left(1-\sum_{j=1}^{P}\Phi_j\beta^{sj}\right)W_t = \left(1-\sum_{j=1}^{q}\theta_j\beta^j\right)\left(1-\sum_{j=1}^{Q}\Theta_j\beta^{sj}\right)a_t \quad (2)$$

Where:

ϕ = Autoregressive parameter,
Φ = Seasonal autoregressive parameter,
θ = Moving average parameter,
Θ = Seasonal moving average parameter,
β = Backward shift operator such that:
$\beta Z_t = Z_{t-1}$For no seasonal observation and
$\beta^s Z_t = Z_{t-s}$, For seasonal observation. Knowing that:
$Z_t = \ln(X_t)$,
X_t= Total number of crashes per month (in this study),
$W_t = \nabla^d \nabla_s^D Z_t$,
∇= Difference operator such that:
$\nabla^d Z_t = Z_t - Z_{t-d}$, For no seasonal effect with an order "d",
$\nabla_s^D Z_t = Z_t - Z_{t-sD}$, For seasonal effect with an order "D"
s = Number of seasons (s=12 for monthly data), and
a_t = the drawing from the distribution of zero mean and constant variance.

The general expression of the SARIMA model in Equation 1, can be expressed in the general order of $(p, d, q) \times (P, D, Q)_s$, where:

p = Order of the autoregressive parameter,
d = Order of difference parameter,
q = Order of moving average parameter,
P = Order of seasonal autoregressive parameter,
D = Order of seasonal difference parameter,
Q = Order of seasonal moving average parameter,
s = Number of seasons ($s = 12$ for monthly data).

For example, if ($p = 0, P = 0, d = 1, D = 1, q = 1, Q = 1$, and $s = 12$), Eq (1) becomes the form shown in Eq (3):

$$W_t = (1-\theta\beta)(1-\Theta\beta^{12})a_t \quad (3)$$

A Box-Jenkins seasonal multiplicative model, which is also called Box-Jenkins seasonal moving average model (SAMIRA), of order (0, 1, 1) × (0, 1, 1)$_{12}$ was used. The order of the model parameters (p, d, q, P, D, Q, s) is suited to properly describe the characteristics of the available data and consequently accurately forecast future data.

The SARIMA $(p, d, q) \times (P, D, Q)_s$ model, which is called Box – Jenkins seasonal model, is to be fitted to a time series through a three-stage procedure. This three-stage procedure is defined as model identification, estimation of model parameters, and diagnostic checking of the estimated parameters.

However, as a start step, all data were plotted against time (time series) to show up important features of the available data, such as trend, seasonality (if any), discontinuities, and outliers. Fig. 1 presents the time series of the total number of crashes in NYC from 2013 to 2019 per month. It can be observed from the figure that there is a little trend, high fluctuation, and seasonal variation for the monthly crash data.

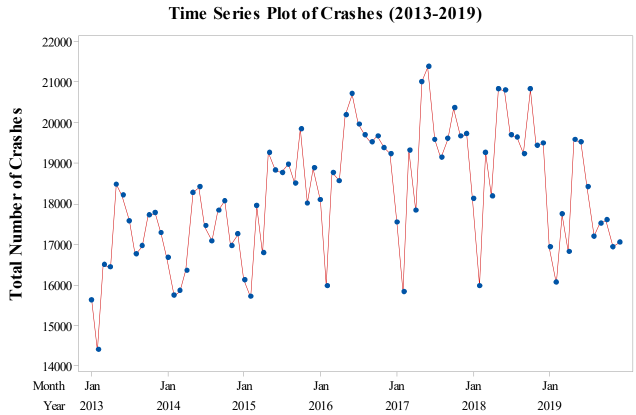


Fig. 1: Time Series Plot for crashes (2013-2019)

In this study, the procedure of fitting the seasonal ARIMA model to the available data is summarized by the following steps [14]:

1. Transforming the data through natural log transformation. This transformation was found to be the most appropriate means to ensure the stationary of the available data.
2. Removing the trend component of the available data by using the first-order differencing ($d = 1$).
3. Removing seasonal variation through the first order seasonal differencing ($D = 1$)
4. Model identification by plotting Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

B. Forecasting Model Identification

Forecasting model identification is to understand the characteristics of the available data to suggest a subclass of models from the general Box-Jenkins family, i.e., Equation 1 for further examination. In other words, identification provides clues about the choice of the Model's parameters order (p, d, q, P, D, Q, s). The order of the seasonal parameter s is 12 because the data is per month every year and there is 12 observation (months) every year. In general, the degrees of differencing d and D are both assumed to be 1, while autocorrelation and partial autocorrelation function are plotted to guess the best orders of p, q, P, and Q. Therefore, the potential model is $(p, 1, q) \times (P, 1, Q)_{12}$. The estimation of the orders of p, P, q and Q can be determined as shown in the model's parameters estimation section (4.2.2).

C. Model's Parameters Estimation

The estimation autocorrelation and partial autocorrelation functions can be used to determine the autoregressive and moving average parameters (p, P, q, and Q). In practice, the unconditional sum of the square method is used to estimate the model parameters. Trial values for the parameters of the $(p, 1, q) \times (P, 1, Q)_{12}$ model are assumed, and the corresponding coefficients of equation (3) are computed (θ and Θ). Then, the sum of squares (SS) is computed for each trial until the minimum sum of squares is obtained.

The number of observations is $N = 84$, $n = N - d - (s \times D)$ =84-1-12= 71 is the number of dependent stochastic components (W_t), d is the degree of simple difference ($d = 1$), D is the degree of seasonal difference ($D = 1$), and s is the length of the periodic cycle ($s = 12$). Both θ and Θ values vary from (-1) to (1). As a start point, it is assumed that $\theta = \Theta = 0.1$. Then with incremental less than 0.001, several iterations are computed to estimate the model's parameters that correspond to the minimum sum of squared errors (SSE). The SSE is the sum of the squared residuals. It quantifies the variation in the data that the ARIMA model does not explain. Table 1 illustrates the iterations to find the minimum SSE and the corresponding parameters (θ and Θ).

Table 1: Estimate of each iteration

Iteration	SSE*	Parameters		
		$\boldsymbol{\theta}$ $(\boldsymbol{q = 1})$	$\boldsymbol{\Theta}$ (Q=1, s=12)	Constant
0	1.17181	0.1	0.1	0.097203
1	0.33029	0.25	0.131	0.034145
2	0.16074	0.4	0.092	0.012329
3	0.10002	0.512	0.242	0.001384
4	0.08603	0.578	0.392	-0.00099
5	0.07736	0.61	0.542	-0.00162
6	0.07174	0.628	0.692	-0.001633
7	0.06798	0.64	0.832	-0.001667
8	0.0679	0.636	0.844	-0.001712
9	0.0679	0.637	0.846	-0.001706
10	0.0679	0.636	0.846	-0.001704

*SSE is the sum of squared error for each iteration.

Table 2 shows the final estimated coefficients of the Model's parameters, the standard error (SE) of the coefficients, the T-value, and the P-value. The standard error of the coefficients estimates the variability between parameter estimates that you would obtain if you took samples from the same population again and again. It is used to measure the precision of the parameter estimate. The smaller the standard error, the more precise the estimate. From Table 2, it can be noticed that the SEs of the coefficients do not exceed 0.2, which is highly acceptable in such kind of estimation. The T-value is just the ratio between the coefficient and its standard error. As can be obtained from Table 2, the P-value is less than 0.001 for all the estimated parameters. Since the p-value is less than 0.05, it can be concluded that the model is statistically significant.

Table 2: Final estimation of parameters

Type	Coefficient	SE Coefficient	T-Value	P-Value
$\theta(q = 1)$	0.6365	0.0959	6.64	<0.001
Θ ($Q = 1, s =$ 12)	0.846	0.112	7.54	<0.001
$Constant$	-0.0017	0.00039	-4.41	<0.001
DF	SS		MS	
68	0.06197		0.00091	

Table 2 summarizes the ARIMA model's degree of freedom (DF), the sum of squares (SS) of residuals, and the mean square error (MS) of the variance of the statistical estimation. The degree of freedom represents the amount of information about the available data. The sum of squares for the residuals is the summation of the residuals using the final parameter estimates, excluding back forecasts. The mean square error is a measure of the accuracy of the fitted model. Smaller values of the mean square error usually indicate a better fitting model. It is recommended to use the MS in the comparison between ARIMA models.

These calculations and tables are repeated with different ARIMA models until the minimum residual squares and mean square error are obtained using the seasonal ARIMA model $(0, 1, 1) \times (0, 1, 1)_{12}$, and the coefficients $\theta = 0.6365$ and $\Theta = 0.846$. Therefore, the model equation is shown below in Eq (4):

$$W_t = (1 - 0.6365\beta)(1 - 0.846\beta^{12})a_t \qquad (4)$$

D. Model Adequacy Verification

Once the model parameters were estimated, the model was checked to decide if it is adequate or not. This can be verified by ensuring that the residuals are independent and normally distributed. Also, the autocorrelation function (ACF) and partial autocorrelation function are used to check the independence of the residuals.

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic is computed as shown in Table 3. This table shows the Lag of the model, chi-square for each lag, degrees of freedom, and the P-value. The lag is defined as the time period that separates the data that are ordered in time. The lags are displayed in multiples of 12. The lag is used to calculate the partial autocorrelation coefficient. The maximum number of lags is approximately $N/4$ for a series with less than 240 observations or $\sqrt{N} + 45$ for a series with more than 240 observations, as recommended by Box and Jenkins. In this particular model, we have 84 observations, so the maximum recommended number of lags is 21. Therefore, the lag of 12 is the most appropriate one and this corresponds to a Chi-square of 5.21, degree of freedom of 9, and P-value of 0.815, as shown in Table 3.

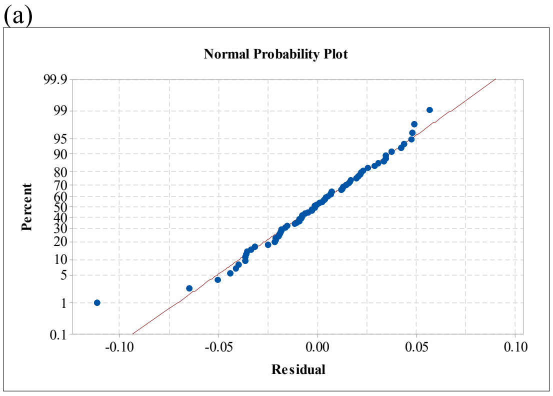
The chi-square value is the test statistic that can be used to determine whether the residuals are independent. It is used to calculate the p-value, which helps to decide whether the residuals are independent. In this particular model, the P-value is 0.815 much greater than 0.05. This is confirmed with 95% confidence that the residuals are independent.

Table 3. Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

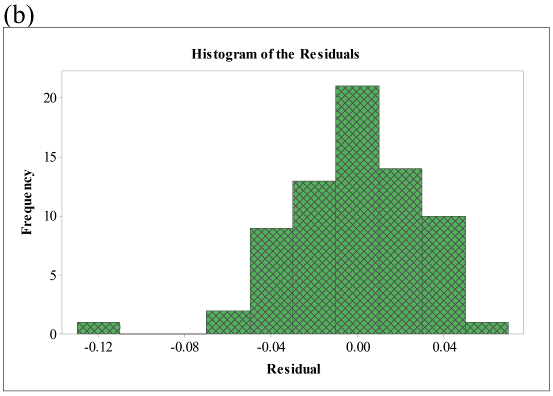
Lag	Chi-Square	DF	P-Value
12	5.21	9	0.815
24	12.23	21	0.933
36	25.26	33	0.830
48	36.35	45	0.818

Other properties of the residuals such as the normality, frequency histogram, residual versus fits, and residual versus order can be observed in Fig. 2. Finally, for further check of the independency of the resulting residuals, the autocorrelation (ACF) and the partial autocorrelation (PACF) of residual series are computed as presented in the figure. The figure shows that most of the computed lags lie inside the tolerance interval ($\pm 2/\sqrt{N}$, at 95% confidence limits). Therefore, the proposed model can be considered an appropriate model due to its capability of removing the dependency from data.

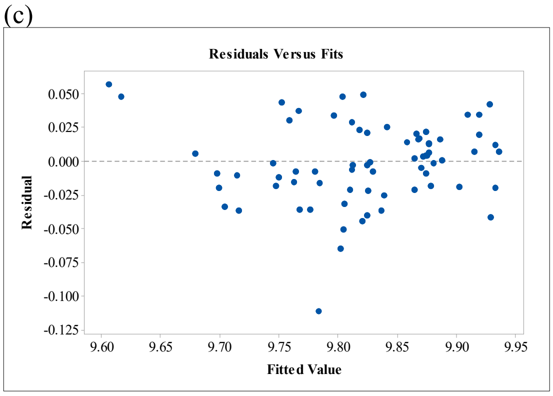
(a)



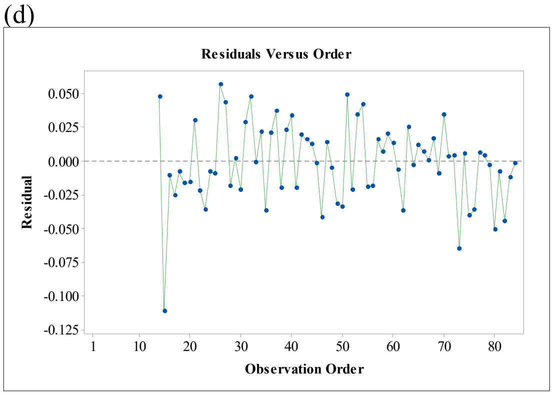
(b)



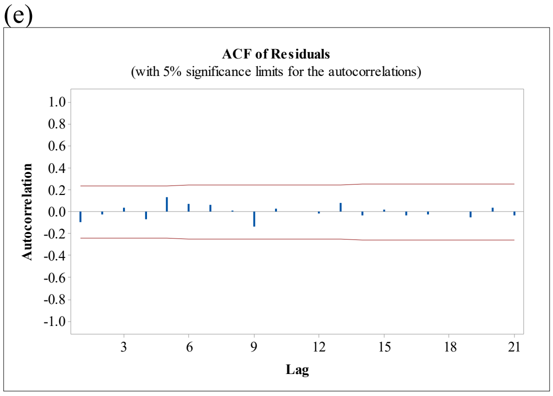
(c)



(d)



(e)



(f)

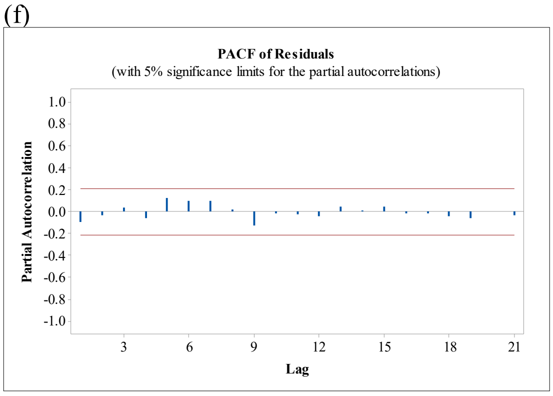


Fig. 2. (a) Normal plot of the residual, (b) Histogram of the residual, (c) Residual versus fitted value, (d) Residual versus observation order, (e) Autocorrelation of residual, and (f) Partial autocorrelation of residual

Eventually, the model is used to forecast the monthly total number of crashes in NYC in the following six years 2020 to 2025. Forecasted data are computed by applying the equation shown below in Eq (5) [19]:

$$\hat{Z}_t(l) = Z_{t+l} = Z_{t+l-1} + Z_{t+l-12} - Z_{t+l-13} + a_{t+l} - \theta a_{t+l-1} - \Theta a_{t+l-12} + \theta\Theta a_{t+l-13} \quad (5)$$

Where t is the origin time ($t = 84$) and l is the lead time (=1, 2, 3, …24). Once the forecasted series Z_t is obtained for ($t = 85, 86, 87 \ldots 108$), then the final estimated series (X_t) is determined by reversing (ln) transformation by taking the exponential of each value in the series of Z_t. To further check the proposed model, it was used to predict the total number of crashes of the last year (2019). Fig. 3 shows the actual total number of crashes per month of the last year (2019) and the total number of crashes obtained from the forecasting model. Both actual and forecasted show good agreement with each other. The corresponding observed values are also presented in Fig. 3. Since there is a good agreement between observed and forecasted values, the model is considered adequate. The total number of crashes per month for the upcoming years 2020 to 2025 is estimated using the proposed model as shown in Fig. 3. For a continuous overview, the time series for the total number of crashes for the years 2019 to 2021 are shown in Fig. 3.

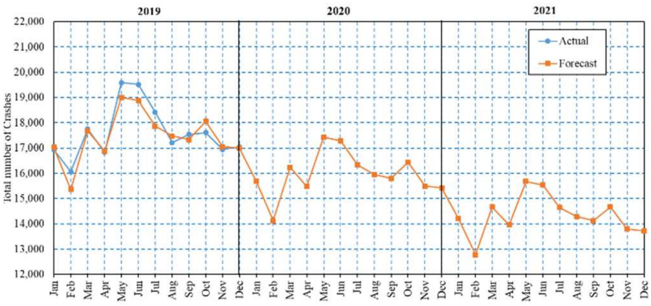


Fig. 3. Actual versus forecasted crashes

Table 4 summarized forecasted values for the total number of crashes per month for the years 2020 to 2025. It can be noticed that the proposed model predicts a noticeable decrease in the total number of crashes in NYC in the future. It can be concluded that the total number of crashes in NYC will continue to decrease from 2020 to 2025. This can be attributed to the successful government traffic safety management that can be observed since the year 2019.

Table 4: Forecast total number of crashes per month for the years from 2020 to 2025

Month/ Year	2020	2021	2022	2023	2024	2025
January	17290	16211	14893	13404	11820	10213
February	15624	14625	13412	12051	10609	9150
March	18015	16834	15412	13824	12149	10461
April	17227	16070	14687	13152	11539	9919
May	19437	18101	16516	14764	12931	11096
June	19337	17977	16375	14613	12777	10946
July	18318	17001	15459	13773	12022	10281
August	17945	16626	15092	13423	11696	9986
September	17813	16476	14931	13256	11532	9828
October	18572	17149	15514	13751	11942	10160
November	17552	16180	14612	12930	11209	9521
December	17518	16121	14534	12839	11111	9422
Total	214647	199370	181438	161781	141337	120982
Average	17887	16614	15120	13482	11778	10082
Standard Deviation	1016	925	835	747	661	580

E. Model Limitations

The ARIMA models are normally suitable for forecasting in short-term periods, but not for forecasting due to the convergence of the autoregressive part of the model to the mean of the time series. The ARIMA model uses the entire data available as input for the model. Therefore, it cannot be used to forecast a specific period in the future e.g., 2030, but it should be a continuous forecasting process on a monthly basis (or similar to the available data). For example, to forecast the total number of crashes in 2030, it is required to forecast 144 observations (12 years × 12 months) using only 84 data available (2013 – 2019). This is not possible and will give unrealistic results. The results depend on the stability of the time series, considering trend and seasonality. If the time series has a lot of randomness, it will be complicated to obtain a confident forecast for the long term. In addition, the forecasting horizon is too extended in comparison with the data available.

IV. CONCLUSION

The crash data for NYC were collected and analyzed. A seasonal ARIMA model was used to represent the data. The model $(0,1,1) \times (0,1,1)_{12}$ was selected, diagnostically checked, and applied to predict the total number of crashes in NYC from January 2020 to December 2025. Also, the proposed model was used for backward prediction of the year 2019 to compare with actual observations. The predicted results showed a good agreement with the actual observed results. The results also showed a strong potential of having a reduction in the total number of crashes in the future. One of the limitations of the analysis conducted in this study is the need for historical crash data, which can be challenging to obtain. Other techniques should be considered such as traffic conflicts analysis [22-25].

V. REFERENCES

[1] W. H. Organization, *Global status report on road safety 2015*. World Health Organization, 2015.

[2] K. Shaaban, A. Siam, and A. Badran, "Analysis of Traffic Crashes and Violations in a Developing Country," *Transportation Research Procedia,* vol. 55, pp. 1689-1695, 2021.

[3] P. Thomas, A. Morris, R. Talbot, and H. Fagerlind, "Identifying the causes of road crashes in Europe," *Annals of advances in automotive medicine,* vol. 57, p. 13, 2013.

[4] K. Shaaban and M. Ibrahim, "Analysis and Identification of Contributing Factors of Traffic Crashes in New York City," *Transportation Research Procedia,* vol. 55, pp. 1696-1703, 2021.

[5] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation research part A: policy and practice,* vol. 44, no. 5, pp. 291-305, 2010.

[6] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Analytic methods in accident research,* vol. 1, pp. 1-22, 2014.

[7] M. S. Ghanim and K. Shaaban, "A Case study for surrogate safety assessment model in predicting real-life conflicts," *Arabian Journal for Science and Engineering,* vol. 44, no. 5, pp. 4225-4231, 2019.

[8] K. Shaaban, I. Gharraie, E. Sacchi, and I. Kim, "Severity analysis of red-light-running-related crashes using structural equation modeling," *Journal of Transportation Safety & Security,* vol. 13, no. 3, pp. 278-297, 2021.

[9] Q. Hussain, W. K. Alhajyaseen, A. Pirdavani, K. Brijs, K. Shaaban, and T. Brijs, "Do detection-based warning strategies improve vehicle yielding behavior at uncontrolled midblock crosswalks?," *Accident Analysis & Prevention,* vol. 157, p. 106166, 2021.

[10] K. Shaaban and A. Pande, "Evaluation of red-light camera enforcement using traffic violations," *Journal of Traffic and*

Transportation Engineering (English Edition), vol. 5, no. 1, pp. 66-72, 2018.

[11] S. Mitra and S. Washington, "On the significance of omitted variables in intersection crash modeling," *Accident Analysis & Prevention,* vol. 49, pp. 439-448, 2012.

[12] X. Yan, E. Radwan, and M. Abdel-Aty, "Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model," *Accident Analysis & Prevention,* vol. 37, no. 6, pp. 983-995, 2005.

[13] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis & Prevention,* vol. 27, no. 3, pp. 371-389, 1995.

[14] A. Montella, L. Colantuoni, and R. Lamberti, "Crash prediction models for rural motorways," *Transportation Research Record,* vol. 2083, no. 1, pp. 180-189, 2008.

[15] C. C. Ihueze and U. O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria," *Accident Analysis & Prevention,* vol. 112, pp. 21-29, 2018.

[16] R. A. Sanusi, F. Adebola, and N. Adegoke, "Cases of road traffic accident in Nigeria: a time series approach," *Mediterranean journal of social sciences,* vol. 7, no. 2 S1, p. 542, 2016.

[17] R. Avuglah, K. Adu-Poku, and E. Harris, "Application of ARIMA models to road traffic accident cases in Ghana," *International journal of statistics and applications,* vol. 4, no. 5, pp. 233-239, 2014.

[18] K. Adu-Poku, R. Avuglah, and E. Harris, "Modeling Road Traffic Fatality Cases in Ghana," *Mathematical Theory and Modelling,* vol. 4, no. 13, pp. 113-120, 2014.

[19] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[20] A. I. McLeod and E. R. Vingilis, "Power computations in time series analyses for traffic safety interventions," *Accident Analysis & Prevention,* vol. 40, no. 3, pp. 1244-1248, 2008.

[21] M. A. Quddus, "Time series count data models: an empirical application to traffic accidents," *Accident Analysis & Prevention,* vol. 40, no. 5, pp. 1732-1741, 2008.

[22] A. Tageldin, T. Sayed, and K. Shaaban, "Comparison of Time-Proximity and Evasive Action Conflict Measures: Case Studies from Five Cities," *Transportation Research Record: Journal of the Transportation Research Board,* no. 2661, pp. 19-29, 2017.

[23] A. Osama, T. Sayed, M. Zaki, and K. Shaaban, "An Inclusive Framework for Automatic Safety Evaluation of Roundabouts," *Journal of Transportation Safety & Security,* vol. 8, no. 4, pp. 377-394, mar 2016, doi: 10.1080/19439962.2015.1107795.

[24] A. Tageldin, T. Sayed, K. Shaaban, and M. Zaki, "Automated Analysis and Validation of Right-Turn Merging Behavior," *Journal of Transportation Safety and Security,* vol. 7, no. 2, pp. 138-152, 2015 2015, doi: 10.1080/19439962.2014.942019.

[25] M. Zaki, T. Sayed, and K. Shaaban, "Use of Drivers' Jerk Profiles in Computer Vision-Based Traffic Safety Evaluations," *Transportation Research Record,* no. 2434, pp. 103-112, 2014 2014, doi: 10.3141/2434-13.