

Article

SARIMA Modelling Approach for Forecasting of Traffic Accidents

Nemanja Deretić ^{1,*}, Dragan Stanimirović ², Mohammed Al Awadh ³, Nikola Vujanović ¹ and Aleksandar Djukić ⁴

¹ Belgrade Business and Arts Academy of Applied Studies, Kraljice Marije 73, 11000 Belgrade, Serbia; nikola.vujanovic@bpa.edu.rs

² Ministry of Transport and Communications of Republic of Srpska, Trg Republike Srpske 1, 78000 Banja Luka, Bosnia and Herzegovina; d.stanimirovic@msv.vladars.net

³ Department of Industrial Engineering, College of Engineering, King Khalid University, P.O. Box 394, Abha 61411, Saudi Arabia; mohalawadh@kku.edu.sa

⁴ Republic Administration for Inspection Affairs of the Republic of Srpska, Trg Republike Srpske 8, 78000 Banja Luka, Bosnia and Herzegovina; a.djukic@inspektorat.vladars.net

* Correspondence: nemanja.deretic@bpa.edu.rs

Abstract: To achieve greater sustainability of the traffic system, the trend of traffic accidents in road traffic was analysed. Injuries from traffic accidents are among the leading factors in the suffering of people around the world. Injuries from road traffic accidents are predicted to be the third leading factor contributing to human deaths. Road traffic accidents have decreased in most countries during the last decade because of the Decade of Action for Road Safety 2011–2020. The main reasons behind the reduction of traffic accidents are improvements in the construction of vehicles and roads, the training and education of drivers, and advances in medical technology and medical care. The primary objective of this paper is to investigate the pattern in the time series of traffic accidents in the city of Belgrade. Time series have been analysed using exploratory data analysis to describe and understand the data, the method of regression and the Box–Jenkins seasonal autoregressive integrated moving average model (SARIMA). The study found that the time series has a pronounced seasonal character. The model presented in the paper has a mean absolute percentage error (MAPE) of 5.22% and can be seen as an indicator that the prognosis is acceptably accurate. The forecasting, in the context of number of a traffic accidents, may be a strategy to achieve different goals such as traffic safety campaigns, traffic safety strategies and action plans to achieve the objectives defined in traffic safety strategies.

Keywords: road traffic; time series; traffic accidents; SARIMA



Citation: Deretić, N.; Stanimirović, D.; Awadh, M.A.; Vujanović, N.; Djukić, A. SARIMA Modelling Approach for Forecasting of Traffic Accidents. *Sustainability* **2022**, *14*, 4403. <https://doi.org/10.3390/su14084403>

Academic Editor: Giovanni Leonardi

Received: 5 March 2022

Accepted: 31 March 2022

Published: 7 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Road transport is a preferred mode of transport due to its low cost and delivery time. General transport and licensing regulations are possible for each country [1]. Transport remains an important development factor in every country [2].

Transport is one of the four existential functions of every living space (work, housing, recreation, and transport), the aim of which is to combine other functions, with as many negative effects as possible. The main harmful consequences of transport today are: depletion of natural resources; pollution of the environment through noise, exhaust fumes and waste materials; traffic accidents (light and serious injury and deaths); property damage; losses; and costs. Active road safety refers to the prevention of road accidents, i.e., reducing the probability of accidents. Active road safety measures reduce the number of road accidents. Passive road safety refers to reducing the harmful consequences of road accidents that have occurred. Society has not always had the same road safety problems (in terms of type and magnitude). These problems did not have the same importance, were not treated in the same way and were not solved in the same way.

Road safety issues are a priority across transport policy (and beyond). It is agreed that road safety problems can be prevented rather than just described and interpreted later. Road safety has found its way into plans for all of transport, alongside urban, economic and other plans. As a result of this new approach to the problem of road safety, incredible results are being achieved, limited to the further development of traffic, with a continuous reduction in the number and consequences of road accidents.

The Road Safety Strategy of the Republic of Serbia [3] for the period from 2015 to 2020 has defined three initial elements for local road safety strategies, namely:

- **Ambition**—To reduce mortality and the risk of serious injury to the level of the most successful countries in the European Union;
- **Mission**—A stable and effective road safety system;
- **Vision**—Road transport without casualties, with a significantly reduced number of injuries and significantly reduced costs of road accidents.

Accepting the recommendations of the United Nations expressed in the document “Global Plan of the Decade of Action for Traffic Safety 2011–2020” prepared by the World Health Organization in different parts of the world [4–6], the strategy identifies five key areas of work (five pillars) to achieve the desired state of traffic safety:

- First pillar: More effective road safety management,
- Second pillar: Safer roads,
- Third pillar: Safer vehicles,
- Fourth pillar: Safer road users and,
- Fifth pillar: Post-crash measures.

In the event of a road accident causing damage to property or injury to people, the driver must stop and inform the competent authority so that the necessary measures can be taken to care for the injured. In order to qualitatively manage road safety, data relevant to road safety and a well-developed database on road safety indicators are needed. The Road Safety Agency of the Republic of Serbia, as one of the most important institutions in the field of road safety, has developed an integrated database on road safety characteristics. In addition to the integrated database on road safety characteristics, all those interested in this field in the Republic of Serbia have access to the portal of publicly available data.

The data will be used to train road accident prediction models to estimate the number of accidents on a monthly basis. The aim of our study is to enable the development of a software tool that will help the municipal services of the City of Belgrade to take preventive measures before major problems arise. The paper uses historical data from the Road Safety Agency and the Ministry of Interior of the Republic of Serbia to predict the future number of road accidents on a monthly basis.

According to the Decade of Action for Road Traffic Safety 2011–2020, it was expected that the number of road traffic accidents would decrease. The main contribution of the study is to show the increase in the number of traffic accidents in the city of Belgrade for the period from 2016 to 2019. Each municipality in the Republic of Serbia allocates certain budgetary funds for strategic documents such as the Road Traffic Safety Strategy and Action Plan for a period of five years.

A particularly important document is the Action Plan, which for each pillar of road safety specifies not only the funds but also the period of the year and the stakeholders who must implement a particular measure. An important contribution of the study is the reference to the last quarter of the year when most traffic accidents occur in the city of Belgrade.

The main motive for the research is to investigate the possibility of applying one of the forecasting models to a time series of traffic accidents in one of the largest cities in the Western Balkans, such as Belgrade.

A particularly important motive was to investigate whether there is a statistically significant period of the year when coordinated activities should be undertaken by the traffic police and all stakeholders to reduce the number of traffic accidents. The main

novelty of the study is that in the last fourth quarter of the year (October, November, and December), most road accidents occurred in the capital of the Republic of Serbia.

The article consists of seven sections. The introductory section focuses on the importance of the theme and justifies why this particular theme was chosen. This is followed by Section 2, which contains a review of the literature. Section 3 contains the materials that make up the data sample used for the study. Section 4 describes the method of data preparation and time series analysis and presents the analysis procedure. Section 5 describes the results of the model obtained. Section 6 then presents and discusses the research findings, while Section 7 presents the conclusions with guidelines for further research studies.

2. Literature Review

The statistical change of many phenomena over time is described by time series. The levels of time series are formed under the influence of a number of long-term and short-term factors as well as a number of random influences.

Due to changes in these influences, there are also fluctuations in the levels of time series that indicate changes in events over time. The annual number of persons killed and injured in road accidents is of great interest in most countries.

Real-time traffic accident detection is very important for drivers and passengers as well as for city services dealing with traffic. Social networks play an important role in the detection of traffic accidents and can help in the analysis [7].

In the UK, Scott [8] analysed the time series of monthly data of road traffic accidents for the period 1970–1978, and Broughton [9] analysed deaths in road traffic accidents for the period 1949–1989. In addition, Quddus [10] studied annual road traffic deaths in the UK between 1950 and 2005. In Sweden, the problem of predicting the number of road traffic deaths was addressed by Brüde [11] based on data from the period 1977–1991. Dadashova et al. [12] used the monthly data on the number of fatal accidents in Spain in the period 2000–2011.

There are many different statistical models used in the analysis of time series of road traffic data. There are two categories of forecasting approaches in transport and traffic: parametric and non-parametric. The main difference is the functional dependence between independent variables and dependent variable [13].

Different approaches have been used to predict changes in traffic accidents: normal linear regression techniques [9], exponential regression techniques [11], Box–Jenkins models and Auto Regressive Integrated Moving Average (ARIMA) [8,10,12]. According to Lavrenz et al. [14], the ARIMA model has been used in many papers modelling time series in road safety research over the last decade [15–19].

In addition to the ARIMA model, Ihueze and Onwurah [20] used the autoregressive integrated moving average with explanatory variables (ARIMAX). The San-gare et al. study [21] shows an approach to predicting road traffic accidents using analytical measures and hybrid machine learning. Almeida et al. [22] propose the Seasonal Auto Regressive Integrated Moving Average (SARIMA) model in their study of traffic flow characteristics.

Artificial neural network algorithms have also been proposed in Refs. [22–28] for the forecasting approach. The most commonly used algorithms are the Feed-Forward Neural Network (FFNN), the Long Short-Term Memory (LSTM), the Convolutional Neural Network (CNN) and a hybrid LSTM-CNN. Naqvi et al. [29] propose SARIMA in their study of the relationship between higher fuel prices and road accidents.

Katrakazas et al. [30] use SARIMA to investigate the impact of COVID-19 on collisions, fatalities and injuries in Greece. Time series trends of the safety effects of pavement resurfacing with SARIMA are shown in the study by Park et al. [31]. Analysis of road crash mortality based on time of occurrence is also performed with SARIMA in the work of Vipin and Rahul [32]. Roland et al. [33] used a multilayer perceptron (MLP) neural network model to predict where and when accidents will occur on a given day and time in the study area. Shannon and Fountas [34] extend the Heston model to predict the collision rate of motor vehicles.

Time Series Regression (TSR) analysis and SARIMA has been applied to assess the relationship between the outcome variable, the number of persons killed due to road traffic accidents, and the variables quantifying the trend and seasonal effects [32,35]. Time-series models include autoregressive integrated moving average (ARIMA), ARIMA with seasonal factors (SARIMA), SARIMA with exogenous explanatory variables (SARIMAX), and nonlinear auto-regression exogenous (NARX) [36]. SARIMA models have been able to adapt and make good predictions even in the presence of anomalies [22]. The SARIMAX model is based on the ARIMA model with the added capability to include seasonality and exogenous parameters [37].

In Ref. [38], several Seasonal Auto-Regressive Integrated Moving Averages with Exogenous factors (SARIMAX) models were used to analyse waste and recycling time-series trends and their relationship with exogenous explanatory variables regarding waste production. In Ref. [39], ARIMA, β ARMA, and KARMA models were compared to forecast the mortality rates due to occupational accidents in the southern region of Brazil. Ref. [40] proposed an ARFIMA-GARCH model for the long memory property in crash risk analysis.

The Autoregressive Integrated Moving Average (ARIMA) is one of the most commonly used forecasting methods for univariate time series data. An extension of ARIMA that supports direct modelling of the seasonal component of the series is called SARIMA or Seasonal ARIMA. In this paper, the SARIMA model is used for short-term forecasts.

When examining the data on the number of traffic accidents in the city of Belgrade, it was found that there is a pronounced seasonal unevenness, which is why the SARIMA model was chosen accordingly.

The main advantages of SARIMA are:

- the model is deterministic and computationally easy;
 - the model has the advantage of requiring multiple model parameters to describe time series that exhibit non-stationarity both within and between seasons;
 - conventional ARIMA cannot capture seasonality and trend in data sets.
- The main disadvantages of SARIMA are:
- the model can only predict a short period of time;
 - the model can only extract linear relationships within the time series data.

3. Materials

Open government data initiatives are on the rise in many countries. The main goals of open access to data are to democratise data access and knowledge production [41]. Open data plays an important role in many applications and services, such as social innovation, policymaking, public opinion research and economic growth [42].

The paper by de Souza et al. [43] highlights the key benefits of Open Government such as transparency, participation and cooperation and uses the term Government 2.0.

The database of open-source data is maintained by government agencies [44]. According to Refs. [44,45], open data can be defined as data produced and funded with public money and made available to the public without restrictions. The data must comply with all privacy and confidentiality laws. For the research, sets of open-source data were adopted from the Republic of Serbia [46,47]. According to the Law on e-Government [48], open data are data that are available for reuse together with metadata in a machine-readable and open form.

According to Ref. [48], data may be re-used by individuals or legal entities for commercial and non-commercial purposes that are different from the original purpose for which they were created.

At the web address data.gov.rs there is the Open Data Portal, where open datasets of the state agencies of the Republic of Serbia are published. In the datasets, there are a number related to road traffic accidents within the topic of public safety. The use cases mention the research of the Data Science Serbia organisation, whose research was conducted on the dataset on traffic accidents in the city of Belgrade for one year (2015).

When searching for the data sets on traffic accidents in road traffic, two types of data sets can be identified:

- Data on traffic accidents for the area of the city of Belgrade [46];
- Data on traffic accidents by police administration and municipality [47].

For these data sets [46,47] each row represents information about a traffic accident.

The first dataset [46] is available in files with the extension .ods (OpenDocument Spreadsheet) for the period from 2015 to 2019 (28 February 2019). In the 2015 dataset, complete data are available for the first 11 months, and only data for one road accident were given for December. As the data for 2015 is not complete, it has not been considered in the analysis. When data for 2019 is considered, only the first two months are available for analysis (January and February). The data for the period from 2016 to 2018 are complete.

For example, the columns in the first dataset for 2016 [46] are:

- Column A: Unique ID road accident number;
- Column B: Date and time at which the accident occurred;
- Column C: Longitude of the place where the road accident occurred;
- Column D: Latitude of the place where the traffic accident occurred;
- Column E: Type of road accident: road accident with damage to property, road accident with injured persons and road accident with fatalities;
- Column F: Name-type of traffic accident: traffic accident with one vehicle, traffic accident with at least two vehicles—without turning, traffic accident with at least two vehicles—turning or crossing, traffic accident with parked vehicles, traffic accident with pedestrians;
- Column G: Detailed description of the traffic accident: traffic accident with one vehicle: 11 types of cases, traffic accident with at least two vehicles—without turning: 9 types of cases, traffic accident with at least two vehicles—turning or crossing: 18 types of cases, traffic accident with parked vehicles: 5 types of cases, pedestrian accident: 25 types of cases.

The second dataset [47] is available in files with the extension .xlsx for the period from 2015 to 2020. The extension.xlsx refers to Microsoft Excel, which has been used in this programme since 2007.

The 2015 dataset is not complete in either source [46,47] and was not used for the study.

For example, the columns in the second dataset for 2016 [47] are:

- Column A: Unique ID road accident number;
- Column B: Police administration;
- Column C: Municipality;
- Column D: Date and time at which the accident occurred;
- Column E: longitude of the place where the traffic accident occurred;
- Column F: latitude of the place where the traffic accident occurred;
- Column G: Type of traffic accident: traffic accident with property damage, traffic accident with injured persons, traffic accident with dead persons;
- Column H: Name-type of traffic accident: traffic accident with one vehicle, traffic accident with at least two vehicles—without turning, traffic accident with at least two vehicles—turning or crossing, traffic accident with parked vehicles, traffic accident with pedestrians;
- Column I: Detailed description of the traffic accident: traffic accident with one vehicle: 11 types of cases, traffic accident with at least two vehicles—without turning: 9 types of cases, traffic accident with at least two vehicles—turning or crossing: 18 types of cases, traffic accident with parked vehicles: 5 types of cases, pedestrian accident: 25 types of cases.

The first dataset [46] and the second dataset [47] have the same columns, but the second dataset has two more columns (police administration and municipality). The number of case types (column G [46] and column I [47]) may differ from the dataset for the year under investigation.

The research area is the city of Belgrade, the capital of the Republic of Serbia. The time frame of the research is the period from 1 January 2016 to 31 December 2019. For the problem studied in this paper, the SARIMA model was used. In this part of the paper, the basic theoretical assumptions of the said model are presented.

The second dataset [47] for the period from 2016 to 2019 was used for the study, with data available for all months of the year (Table 1). Open data were used in the work, representing the effort of the Republic of Serbia in partnership with the United Nations Development Programme to gain new values through research projects.

Table 1. Number of traffic accidents by month for the city of Belgrade (2016–2019).

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2016	1282	1220	1496	1471	1446	1429	1255	1290	1499	1640	1533	1727
2017	1380	1129	1480	1472	1516	1543	1334	1304	1591	1726	1723	1776
2018	1410	1332	1618	1540	1542	1415	1400	1399	1487	1733	1495	1705
2019	1529	1310	1529	1426	1500	1451	1414	1299	1408	1627	1565	1659

Data on the number of traffic accidents by month from 2016 to 2018 (36 months) were used to build the model, while data on the number of traffic accidents by month from 2019 (12 months) were used to test the accuracy of the model.

Data on the number of road accidents by month from 2020 were not used due to the circumstances caused by the COVID-19 pandemic.

Limitations of the study are:

- The analysis of road accidents included only columns with latitude and longitude from the data sets;
- The time series was only examined on a monthly basis;
- Only years with complete data were used;
- The year 2020 was not included in the analysis due to conditions caused by the COVID-19 pandemic;
- Data were available for 48 months, with a model developed based on 36 months and a model that was tested based on 12 months (2019);
- Data are limited due to constraints imposed by public use, so data on gender, age, length of driving experience and other driver characteristics are not available, but neither are data on the vehicle(s) involved in the accident.

4. Methodology

The advantages of SARIMA lie in its well-known statistical properties and its effective modelling process. The SARIMA model is one of the most effective linear models for seasonal time series forecasting. Although SARIMA is not a new method, it has been applied in this work in an innovative way, which implies its application under new conditions. In this paper, SARIMA was implemented using common statistical software, such as the R programming language and RStudio. The main advantage of SARIMA processes is its ability to model time series with trends, seasonal patterns and short-term correlation with a small data set.

The following steps are followed in the application of SARIMA time series analysis [49]:

- Decomposition of the time series,
- Autocorrelation and partial autocorrelation,
- Stationarity test,
- SARIMA modelling,
- Residual test and test set error,
- Prediction.

Box et al., [50] introduced the ARIMA model. Since seasonal differentiation was required to make seasonal time series stationary, the SARIMA model was introduced. The SARIMA model has four components [19,50]:

- The non-seasonal and seasonal Auto Regressive (AR) polynomial term of order p and P , Equations (1) and (2):

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (1)$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (2)$$

- The non-seasonal and seasonal Moving Average (MA) part of order q and Q , Equations (3) and (4):

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (3)$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (4)$$

- Non-seasonal differencing operator is the of order d used to eliminate polynomial trends, Equation (5):

$$(1 - B)^d \quad (5)$$

- Seasonal differencing operator is the order of D used to eliminate seasonal patterns, Equation (6):

$$(1 - B^s)^D \quad (6)$$

Parameters ϕ and θ are the ordinary ARMA coefficients, Φ and Θ are the seasonal ARMA coefficients, B is the backshift operator, whose effect on a time series Y_t can be summarized as Equation (7):

$$B^d Y_t = Y_{t-d} \quad (7)$$

Therefore, the generalized form of the $SARIMA(p, d, q) \times (P, D, Q)_s$ model for a series Y_t can be written as in Ref. [50], Equation (8):

$$\phi_p(B)\Phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (8)$$

where s is the length of the periodicity (seasonality) and ε_t is a white noise sequence.

The following notations were used in the formulas for obtaining the value of the forecasting errors:

- y_t —actual values;
- f_t —forecast values;
- $e_t = y_t - f_t$ —forecast error.

Accuracy of the model was computed through three measures:

- Mean Absolute Error (MAE) or Mean Absolute Deviation (MAD), Equation (9):

$$MAE(MAD) = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (9)$$

- Mean Absolute Percentage Error (MAPE), Equation (10):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \times 100 \quad (10)$$

- Theil's U_1 —statistics is a measure of forecast accuracy ($0 \leq U \leq 1$; $U = 0$ means a perfect fit), Equation (11):

$$U = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} / \left(\sqrt{\frac{1}{n} \sum_{t=1}^n y_t^2} + \sqrt{\frac{1}{n} \sum_{t=1}^n f_t^2} \right) \quad (11)$$

5. Results

In this section, the results of the prediction with the SARIMA method are presented. The R programming language was used for the analysis. The programming language R is a powerful modern computing environment for data manipulation, statistical calculations and visualisation [51].

5.1. Basic Data on the Time Series

Figure 1 below shows the time series used in the study. It shows that the number of traffic accidents has a slightly positive growth trend. The average number of traffic accidents per year is 1441 (2016), 1498 (2017), 1506 (2018) and 1476 (2019).

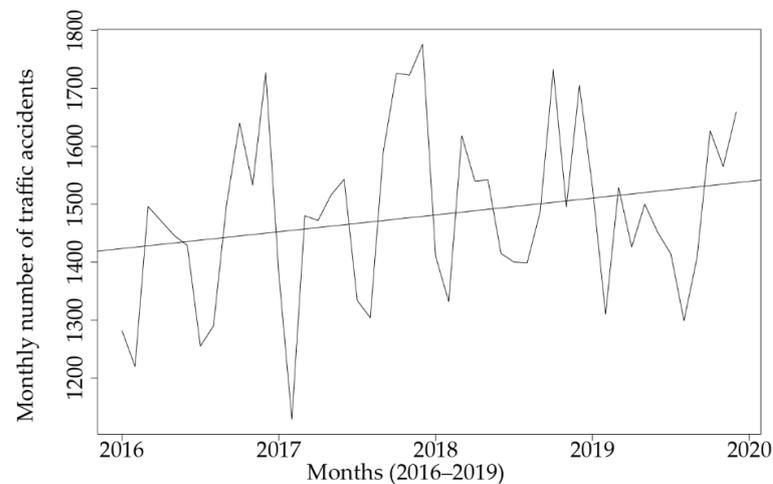


Figure 1. Time series with data on traffic accidents by months (2016–2019).

Taking 2016 as the base year, we can see that the average number of road accidents increased by 3.97% (2017), 4.56% (2018) and 2.48% (2019).

The basic descriptive data on the time series can be seen from the box-plot diagram in Figure 2 below.

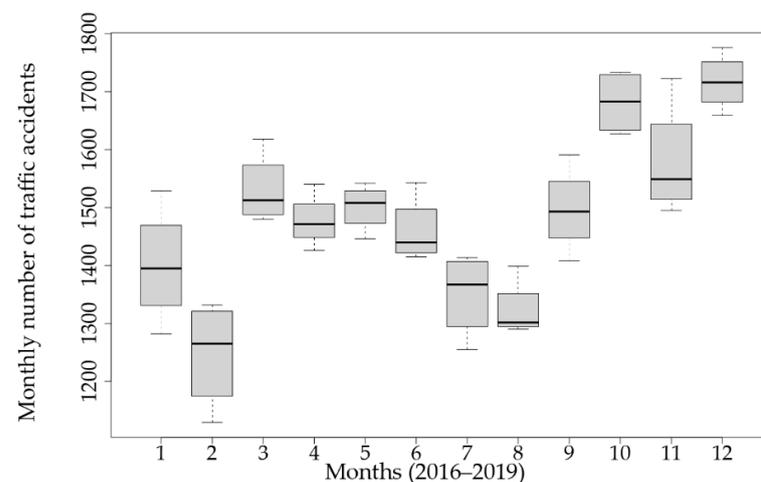


Figure 2. Box-plot diagram with data on traffic accidents by months (2016–2019).

The lowest number of road accidents occurs in February, when the average number of accidents is 1248 for the four-year period studied. Looking at the quarters during the year,

the first and third quarters correspond to approximately 1393 and 1390 road accidents on a monthly basis. The most dangerous time of the year for drivers, passengers, pedestrians and other road users is certainly the fourth quarter, when an average of 1659 reported road accidents occur per month. The most unsafe month of the year is December, which logically belongs to the fourth quarter, when an average of 1717 traffic accidents occur per month. The ratio between the maximum average value for the month (December) and the minimum average value for the month (February) is 37.59%.

The standard deviations of the monthly number of road accidents by year are: 150 (2016), 184 (2017), 121 (2018) and 108 (2019). If quarters are observed during the year, then the standard deviations of the monthly number of traffic accidents are the same: 74 (quarter 1), 42 (quarter 2), 57 (quarter 3) and 59 (quarter 4).

5.2. Development of the SARIMA Model

In order to use the SARIMA model, a certain data arrangement is required. Data from 2016 to 2018 were used to build the model, while data from 2019 are used for testing. The following figure shows the logarithmic values of the number of traffic accidents by month (Figure 3), and the next figure (Figure 4) shows $\text{diff}(\log(\text{values}))$. The following libraries of the R programming language were used for data analysis: MASS [52], tseries [53], forecast [54] and astsa [55].

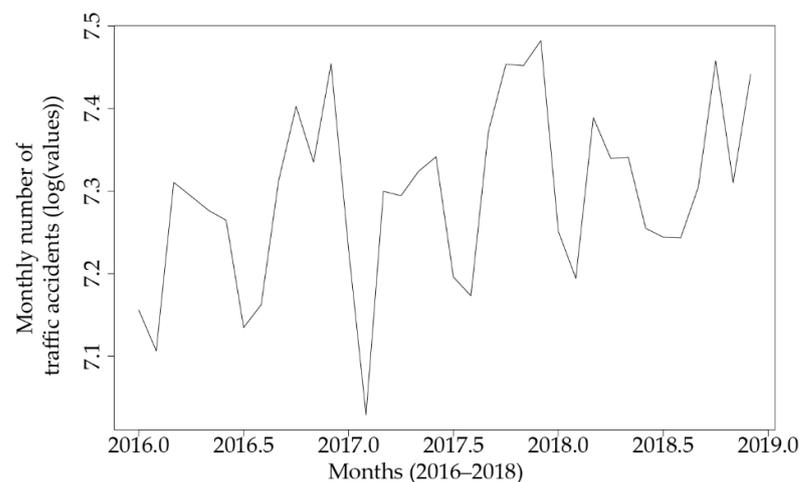


Figure 3. log values of the number of traffic accidents by months (2016–2018).

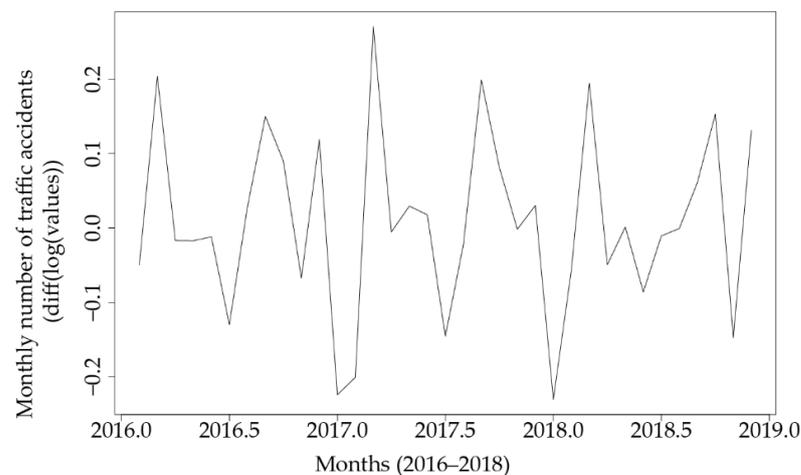


Figure 4. diff (log) values of the number of traffic accidents by month (2016–2018).

The analysis and modelling of the SARIMA model is based on the Box–Jenkins method, which comprises three steps:

- Model identification;
- Model estimation;
- Model validation.

According to Ref. [56], many variables are used in logarithms (logs) for forecasts and economic analyses. In the analysis of time series, this transformation is often used to stabilise the variance of a series.

The application of the function `auto.arima` with seasonal influence in the programming language R (Table 2) showed that it is the best model $ARIMA(0,1,2) \times (1,1,0)^{12}$. In Table 2 is presented information criteria such as the Akaike information criterion (AIC), the Akaike information criterion with a correction for small sample sizes (AICc) and the Bayesian information criterion (BIC).

Table 2. Results of the `auto.arima` function.

Series: log(accidents)			
ARIMA (0,1,2) × (1,1,0) ¹² (2016–2018)			
Coefficients:			
	ma1	ma2	sar1
	−0.7383	0.1002	−0.7593
s.e.	0.2315	0.2232	0.1257
sigma ² estimated as 0.002289		log likelihood = 33.6	
AIC = −59.19	AICc = −56.97		BIC = −54.65

In statistics, AIC is used to compare different possible models and determine which one is the best fit for the data. The basic principles for the use of AIC are:

- A lower value indicates a simpler model compared with a model with a higher AIC;
- It is a relative measure of model parsimony, i.e., it is only significant when we compare the AIC for alternative hypotheses (=different models of the data).

The information about the AICc value of the model (the lower case ‘c’ indicates that the value was calculated from the AIC test corrected for small sample sizes). The smaller the AIC value, the better the model fits.

The Bayesian Information Criterion (BIC) is a criterion for model selection from a finite set of models. It is based in part on the likelihood function and is closely related to the Akaike information criterion (AIC). As the complexity of the model increases, the BIC value increases and as the likelihood increases, the BIC value decreases (a lower value is better).

The test results of the SARIMA model can be found in Table 3, based on the actual values and the predicted values of 2019.

Based on Equations (9)–(11), the calculated prediction error indicators are calculated in Table 4. The mean absolute error, MAE = 77, is the average (absolute) difference between the actual value and the forecast value for 2019. The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of the forecasts.

The MAPE is a relative measure that expresses the errors as a percentage of the actual data. This is its greatest advantage, as it provides a simple and intuitive way to assess the magnitude or significance of errors. Theil’s U1 statistic ranges from 0 to 1, with values closer to 0 indicating higher predictive accuracy.

The interpretation of typical MAPE values is shown in Table 5, according to Ref. [57].

Table 3. Results of error estimate for 2019.

Month.Year	y_t	f_t	y_t^2	f_t^2	e_t	e_t^2	$ e_t $	$ e_t/y_t $
1.2019	1529	1379	2337,841	1901,641	150	22,500	150	0.0981
2.2019	1310	1165	1716,100	1357,225	145	21,025	145	0.1107
3.2019	1529	1499	2337,841	2247,001	30	900	30	0.0196
4.2019	1426	1475	2033,476	2175,625	−49	2401	49	0.0344
5.2019	1500	1509	2250,000	2277,081	−9	81	9	0.0060
6.2019	1451	1498	2105,401	2244,004	−47	2209	47	0.0324
7.2019	1414	1338	1999,396	1790,244	76	5776	76	0.0537
8.2019	1299	1315	1687,401	1729,225	−16	256	16	0.0123
9.2019	1408	1552	1982,464	2408,704	−144	20,736	144	0.1023
10.2019	1627	1712	2647,129	2930,944	−85	7225	85	0.0522
11.2019	1565	1650	2449,225	2722,500	−85	7225	85	0.0543
12.2019	1659	1743	2752,281	3038,049	−84	7056	84	0.0506
Total	17,717	17,835	26,298,555	26,822,243	−118	97,390	920	0.6266

Table 4. Calculated forecasting error indicators.

Model	MAE/MAD	MAPE	Theil’s U1 Statistics
$SARIMA(0,1,2) \times (1,1,0)^{12}$	77	5.22%	0.0303

Table 5. Interpretation of typical MAPE values.

MAPE	Interpretation
<10	Highly accurate forecasting
10–20	Good forecasting
20–50	Reasonable forecasting
>50	Inaccurate forecasting

MAPE is commonly used because it is easy to interpret and explain. For example, a MAPE value of 5.22% means that the average difference between the forecasted value and the actual value is 5.22%. It measures this accuracy as a percentage which can be calculated as the average absolute percent error for each time period minus actual values divided by actual values.

The logarithmic values for the prediction of the number of road accidents by month with confidence intervals of 80% and 95% are shown in Figure 5.

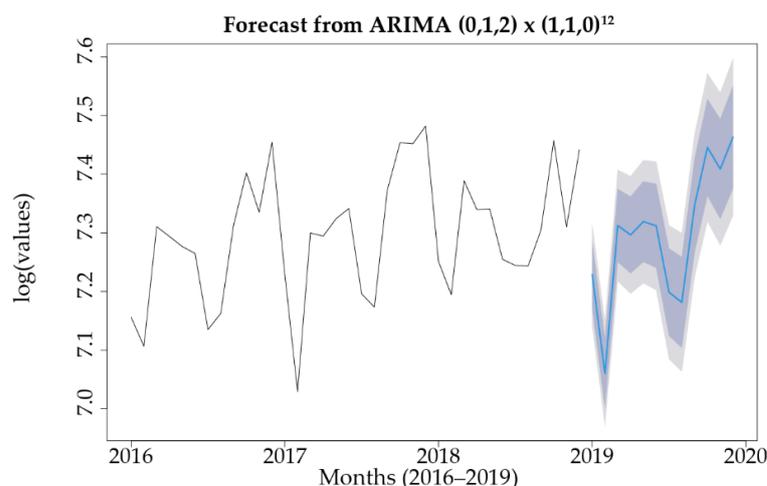


Figure 5. Log of forecasting values and confidence intervals for 2019.

Log transformation is one of the most popular data transformation techniques. It is mainly used to transform a skewed distribution into a normal/less skewed distribution. In other words, the log transformation reduces or eliminates the skewness of our original data.

The values of the forecast of the number of traffic accidents by month for 2019 (dashed line) are shown in Figure 6.

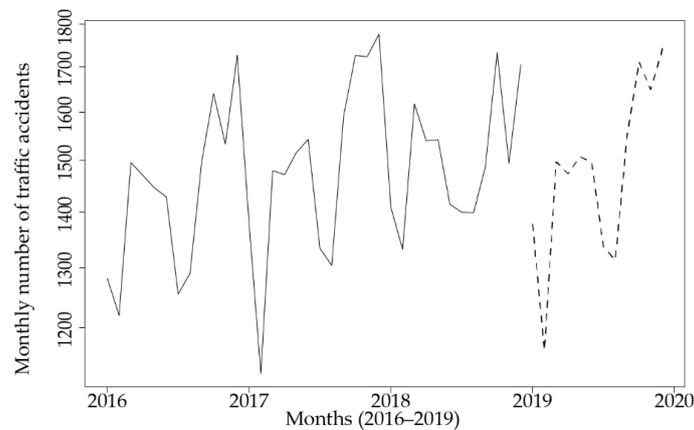


Figure 6. Forecasted values for 2019.

Figure 6 shows the pattern of traffic accidents, which is repeated from year to year with small changes. It can be seen that after a lower number of accidents in the first quarter, the number of accidents increases in the second quarter.

Due to the holiday season, the number of road accidents decreased in the third quarter. In the period studied (2016–2019), most traffic accidents occurred in the last month of the third quarter (September—the beginning of the school year for primary and secondary schools) and in the last quarter.

6. Discussion

According to Almeida et al. [22], the SARIMA method is used for short-term forecasts. In this work, 75% of the data is used for training (data from 2016–2018) and another 25% for testing (data from 2019).

The choice of the best SARIMA model is shown in Table 6.

Table 6. Choice of best SARIMA model—Data (2016–2018).

SARIMA	AIC
SARIMA (2,1,2) × (1,1,1) ¹²	Inf
SARIMA (0,1,0) × (0,1,0) ¹²	−46.66586
SARIMA (1,1,0) × (1,1,0) ¹²	Inf
SARIMA (0,1,1) × (0,1,1) ¹²	Inf
SARIMA (0,1,0) × (1,1,0) ¹²	Inf
SARIMA (0,1,0) × (0,1,1) ¹²	Inf
SARIMA (0,1,0) × (1,1,1) ¹²	Inf
SARIMA (1,1,0) × (0,1,0) ¹²	−50.33068
SARIMA (1,1,0) × (0,1,1) ¹²	Inf
SARIMA (1,1,0) × (1,1,1) ¹²	Inf
SARIMA (2,1,0) × (0,1,0) ¹²	−49.97647
SARIMA (1,1,1) × (0,1,0) ¹²	−51.32873
SARIMA (1,1,1) × (1,1,0) ¹²	−59.1354
SARIMA (1,1,1) × (1,1,1) ¹²	Inf
SARIMA (1,1,1) × (0,1,1) ¹²	Inf
SARIMA (0,1,1) × (1,1,0) ¹²	Inf
SARIMA (2,1,1) × (1,1,0) ¹²	−57.37561
SARIMA (1,1,2) × (1,1,0) ¹²	Inf
SARIMA (0,1,2) × (1,1,0) ¹²	−59.19096
SARIMA (0,1,2) × (0,1,0) ¹²	−51.33234
SARIMA (0,1,2) × (1,1,1) ¹²	Inf
SARIMA (0,1,2) × (0,1,1) ¹²	Inf
SARIMA (0,1,3) × (1,1,0) ¹²	−57.33942
SARIMA (1,1,3) × (1,1,0) ¹²	Inf
Best model:	SARIMA (0,1,2)(1,1,0) [12]

Since a low AIC value indicates a simpler model compared with a model with a higher AIC value, SARIMA (0,1,2) × (1,1,0)¹² was selected as the best fitting model.

The model identification is based on a comparison of the autocorrelation functions (ACF) of the partial autocorrelation (PACF) with the theoretical profiles of these functions.

Model identification is characterised by considerable subjectivity. To minimise subjectivity and improve the process of determining the ranks of the ARIMA process, some of the model selection criteria are used.

The best known are information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) and the normalised version of the BIC.

All criteria are based on the assessment of the fit of nonlinear models, taking into account the number of model parameters.

They consist of the natural logarithm of the least square error and the penalty for the number of estimated parameters [58,59], Equations (12)–(14):

$$AIC = T \ln(MSE) + 2k \quad (12)$$

$$BIC = T \ln(MSE) + k \ln(T) \quad (13)$$

$$Normalized\ BIC = \ln(MSE) + k \frac{\ln(T)}{T} \quad (14)$$

In Equations (12)–(14), T represents the number of observations and k the number of parameters of the model ($k = p + q + P + Q + 1$). The number MSE is the mean square error. The result of the identification steps is represented by the corresponding model structure $(p,d,q) \times (P,D,Q)^S$. Model estimation is based on fitting the model selected in the previous step and determining the model parameters. This step is based on the non-linear least squares and maximum reliability methods.

Model validation, the final step of the Box–Jenkins method, involves analysing the stationarity, invertibility and redundancy of the model parameters [60]. If the residuals, i.e., the difference between the actual values and the values estimated by the model, are random, the model is satisfactory. Otherwise, it is necessary to repeat the process of model identification and estimation and find a better model.

The following figures (Figures 7 and 8) show the plot of the autocorrelation function (ACF) and the partial ACF (PACF).

Our aim now is to find a suitable SARIMA model based on the ACF and PACF shown in Figures 7 and 8 for the 2016–2018 data, in order to test it with the 2019 data. The significant swings at Lag 1, 7 and 13 in the ACF indicate a non-seasonal MA (1) component, and the significant swings at Lag 5 and 9 in the ACF indicate a seasonal MA (1) component.

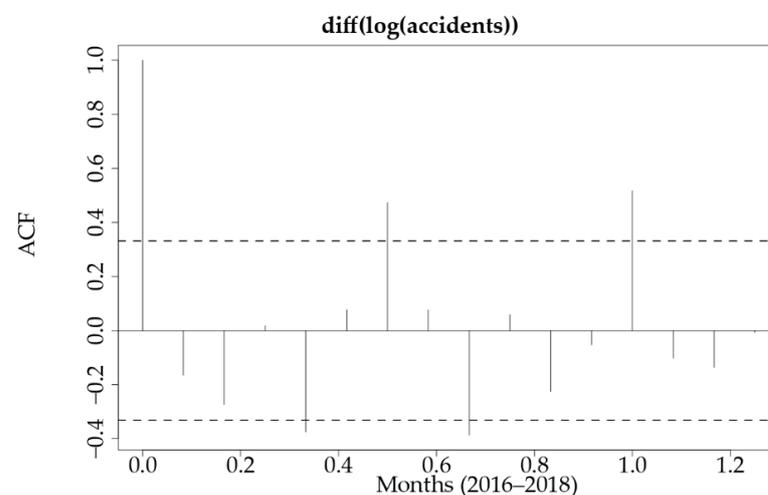


Figure 7. ACF diagram (2016–2018).

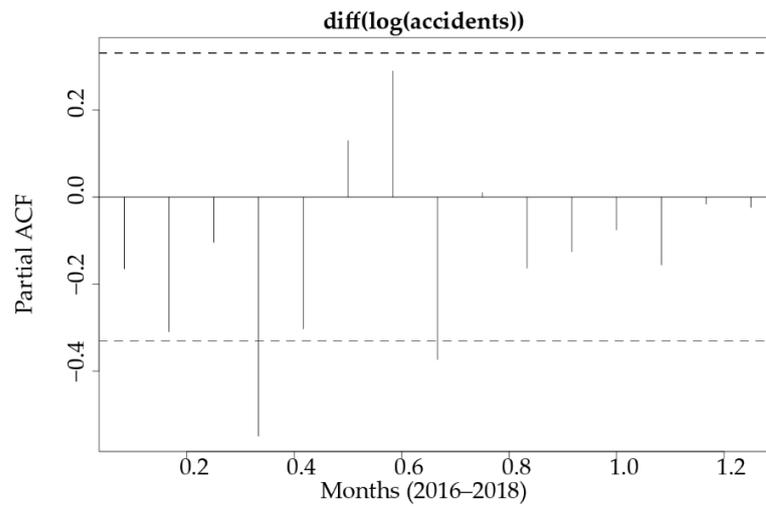


Figure 8. PACF diagram (2016–2018).

The PACF show significant swings at lag 4 and 8, suggesting that some additional non-seasonal terms need to be included in the model. We also tried other models with AR terms, but none yielded a smaller AICc value. Therefore, we decided to use the SARI-MA $(0,1,2) \times (1,1,0)^{12}$ model.

When all the data (2016–2019) were used to build the model, applying the auto.arima function resulted in the following model SARIMA $(0,1,1) \times (1,1,0)^{12}$. The results of the residual analysis are shown in Figure 9. Almost all spikes are now within the significance limits in the ACF diagram.

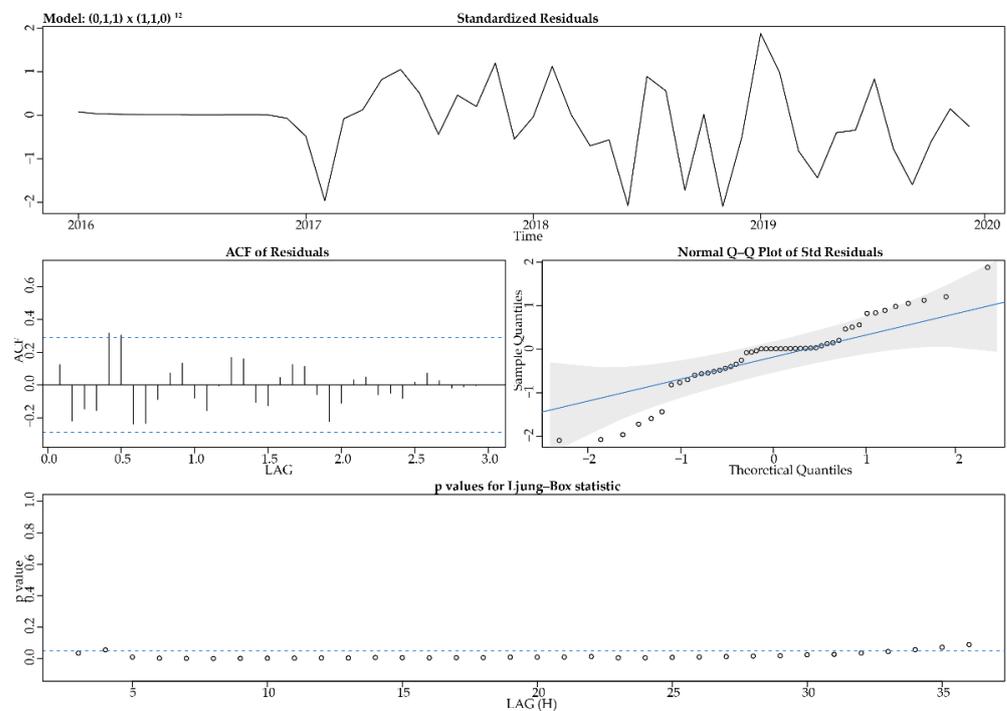


Figure 9. Result analysis results of the SARIMA model $(0,1,1) \times (1,1,0)^{12}$ (2016–2019).

The standardised residual is determined by dividing the difference between the observed and expected values by the square root of the expected value. One type of residual that we often use to identify outliers in a model is the standardised residual. From the figures of the standardised residuals and the ACF of the residuals, we can see that there are

no large outliers. The figure Normal Q-Q plot of Std residuals shows that only 5 of the 48 values are outside the confidence interval.

A significant p -value in the Ljung–Box statistical test rejects the null hypothesis that the time series is not autocorrelated. The figure of p -values for the Ljung–Box statistic shows that there is no basis for rejecting the hypothesis.

To show the robustness of the proposed model, a comparison is made with other known methods from Ref. [61]. The dataset from Ref. [61] corresponds to the monthly production of electrical equipment (computers, electronic and optical products) in the Eurozone (17 countries) in the period January 1996–March 2012. In the study from Ref. [61], it can be seen that the SARIMA method is in third place among the ten selected methods (with the lowest value) and the MAPE value is 0.09% higher compared with the first two methods. Ten selected methods were (in decreasing order for value MAPE): Naïve, Prophet, Seasonal Naïve, NNETAR, Exponential smoothing, ARIMA + Decomposition, Exponential smoothing + Decomposition, SARIMA, GARCH + Decomposition and TBATS. [61].

7. Conclusions

As in other studies [62–64], the number of traffic accidents was considered on a monthly basis. The study of spatio-temporal correlation of traffic accidents is addressed in papers [63,64] that are not included in our paper and represent one of the directions for further research. The main difference with the work studied is that this work clearly indicates that most traffic accidents occur in the last quarter of the year. If the spatio-temporal correlation was undertaken on the basis of individual zones and different years, then we would see the results of the work on road safety prevention. This paper presents an analysis of the time series that gave a clear picture of the existence of a pattern in the number of road accidents on a monthly basis for the period 2016–2019.

Predicting the number of traffic accidents during certain periods of the year is an important part of traffic management in a given area. Informing citizens about the status of road safety contributes significantly to understanding the problem, improving citizens' attitudes towards road safety and improving traffic behaviour. Assistance and support to other road safety sectors should contribute to their work and to the improvement of certain aspects of road safety.

Improving the monthly forecast of the number of traffic accidents can have several positive effects on the city and its citizens. For example, identifying problematic months of the year can lead the relevant authorities to take action to combat the problems through changes to the road safety strategy.

The results of the analysis of the time series and the SARIMA method show that for four consecutive years, most road accidents occurred in the fourth quarter. The road safety campaign is a system of activities whose general objective is to promote safer road use. The specific objectives of road safety campaigns relate to changing road knowledge, attitudes, skills and behaviour, all with the aim of improving road safety.

In this paper, a mathematical model for traffic accident prediction was proposed for the case of the city of Belgrade. The result of the model should be taken with caution, as not all road accidents are reported to the police, especially those involving minor property damage.

The results of applying the libraries of the R programming language show that the SARIMA $(0,1,2) \times (1,1,0)^{12}$ model is the most suitable for modelling according to the open data from 2016 to 2018. If open data from 2016 to 2019 were used to build the model, then the best model is SARIMA $(0,1,1) \times (1,1,0)^{12}$. Future research will focus on developing forecasting models for other municipalities in the Republic of Serbia as well as for other countries.

The study of predicting the number of traffic accidents on a monthly basis has not yet been fully explored. In the future, we plan to test new algorithms to detect anomalies (holidays, specific routes, etc.). This research can be undertaken through a hybrid method composed of statistical and neural network approaches presented in numerous literature.

Modelling and prediction can be further improved to extract more desirable data on traffic accidents and reduce complexity.

As motorisation increases, so does the number of traffic accidents and all the negative consequences associated with them. Therefore, a constant review, analysis and evaluation of the existing situation is necessary. Statistically, there are about 47 road accidents with injured persons and about 130 road accidents with property damage for every road accident with fatalities, as shown in the available data for 2016.

This paper has shown that meaningful results and conclusions can be obtained from open data from the database on traffic accidents. These results and conclusions can be used in determining preventive measures and campaigns for road safety in the city of Belgrade.

A possible measure and improvement in the management of open data on road accidents could be to increase the amount of information published on the open data portal. Of course, care should be taken to protect the personal data of road accident participants.

Some general data can be published in the open data from the investigation documentation, such as the meteorological conditions prevailing at the time of the accident or data on general visibility.

Among other things, it is necessary to provide information on the general condition of the vehicle when the vehicle is referred for a technical investigation (e.g., condition of the brake system).

A special category could refer to the driver and general information about them (e.g., age, profession, length of driving experience, whether he is a weekend driver, whether they were under the influence of alcohol or opiates). Finally, a column could be added to the note containing information that the investigating agencies consider important without being included in any of the previous columns.

Author Contributions: Conceptualization, N.D. and D.S.; methodology, N.D. and N.V.; validation, M.A.A. and A.D.; formal analysis, N.D.; writing—original draft preparation, N.D., N.V. and A.D.; writing—review and editing, D.S. and M.A.A.; All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research, King Khalid University for funding this work through the General Research Project under grant number (RGP2/163/43).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Memiş, S.; Demir, E.; Karamaşa, Ç.; Korucuk, S. Prioritization of road transportation risks: An application in Giresun province. *Oper. Res. Eng. Sci. Theory Appl.* **2020**, *3*, 111–126. [[CrossRef](#)]
- Sénquiz-Díaz, C. Transport infrastructure quality and logistics performance in exports. *ECONOMICS-Innov. Econ. Res.* **2021**, *9*, 107–124. [[CrossRef](#)]
- Pešić, D.; Pešić, A. Monitoring of Road Safety Performance Indicators—Current Situation and Trends in The Republic of Serbia. *Transp. Res. Procedia* **2020**, *45*, 70–77. [[CrossRef](#)]
- Pérez-Núñez, R.; Hidalgo-Solórzano, E.; Híjar, M. Impact of Mexican Road Safety Strategies implemented in the context of the UN's Decade of Action. *Accid. Anal. Prev.* **2021**, *159*, 106227. [[CrossRef](#)]
- Bliss, T.; Breen, J. Meeting the management challenges of the Decade of Action for Road Safety. *IATSS Res.* **2012**, *35*, 48–55. [[CrossRef](#)]
- Morimoto, A.; Wang, A.; Kitano, N. A conceptual framework for road traffic safety considering differences in traffic culture through international comparison. *IATSS Res.* **2021**, *46*, 3–13. [[CrossRef](#)]
- Ali, F.; Ali, A.; Imran, M.; Naqvi, R.A.; Siddiqi, M.H.; Kwak, K.S. Traffic accident detection and condition analysis based on social networking data. *Accid. Anal. Prev.* **2021**, *151*, 105973. [[CrossRef](#)]
- Scott, P.P. Modelling time-series of British road accident data. *Accid. Anal. Prev.* **1986**, *18*, 109–117. [[CrossRef](#)]
- Broughton, J. Forecasting road accident casualties in Great Britain. *Accid. Anal. Prev.* **1991**, *23*, 353–362. [[CrossRef](#)]

10. Quddus, M.A. Time series count data models: An empirical application to traffic accidents. *Accid. Anal. Prev.* **2008**, *40*, 1732–1741. [[CrossRef](#)]
11. Brüde, U. What is happening to the number of fatalities in road accidents? A model for forecasts and continuous monitoring of development up to the year 2000. *Accid. Anal. Prev.* **1995**, *27*, 405–410. [[CrossRef](#)]
12. Dadashova, B.; Arenas-Ramírez, B.; Mira-McWilliams, J.; Aparicio-Izquierdo, F. Methodological development for selection of significant predictors explaining fatal road accidents. *Accid. Anal. Prev.* **2016**, *90*, 82–94. [[CrossRef](#)]
13. Milenković, M.; Švadlenka, L.; Melichar, V.; Bojović, N.; Avramović, Z. SARIMA modelling approach for railway passenger flow forecasting. *Transport* **2018**, *33*, 1113–1120. [[CrossRef](#)]
14. Lavrenz, S.M.; Vlahogianni, E.I.; Gkritza, K.; Ke, Y. Time series modeling in traffic safety research. *Accid. Anal. Prev.* **2018**, *117*, 368–380. [[CrossRef](#)]
15. Carnis, L.; Blais, E. An assessment of the safety effects of the French speed camera program. *Accid. Anal. Prev.* **2013**, *51*, 301–309. [[CrossRef](#)]
16. Commandeur, J.J.; Bijleveld, F.D.; Bergel-Hayat, R.; Antoniou, C.; Yannis, G.; Papadimitriou, E. On statistical inference in time series analysis of the evolution of road safety. *Accid. Anal. Prev.* **2013**, *60*, 424–434. [[CrossRef](#)]
17. Quddus, M.A. Non-Gaussian interrupted time series regression analysis for evaluating the effect of smart motorways on road traffic accidents (No. 16–0157). In Proceedings of the TRB Annual Meeting, Washington, DC, USA, 10–14 January 2016.
18. Sebegu, M.; Naumann, R.B.; Rudd, R.A.; Voetsch, K.; Dellinger, A.M.; Ndlovu, C. The impact of alcohol and road traffic policies on crash rates in Botswana, 2004–2011: A time-series analysis. *Accid. Anal. Prev.* **2014**, *70*, 33–39. [[CrossRef](#)]
19. Vanlaar, W.; Robertson, R.; Marcoux, K. An evaluation of Winnipeg’s photo enforcement safety program: Results of time series analyses and an intersection camera experiment. *Accid. Anal. Prev.* **2014**, *62*, 238–247. [[CrossRef](#)]
20. Ihueze, C.C.; Onwurah, U.O. Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. *Accid. Anal. Prev.* **2018**, *112*, 21–29. [[CrossRef](#)]
21. Sangare, M.; Gupta, S.; Bouzefrane, S.; Banerjee, S.; Muhlethaler, P. Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. *Expert Syst. Appl.* **2020**, *167*, 113855. [[CrossRef](#)]
22. Almeida, A.; Brás, S.; Oliveira, I.; Sargento, S. Vehicular traffic flow prediction using deployed traffic counters in a city. *Future Gener. Comput. Syst.* **2022**, *128*, 429–442. [[CrossRef](#)]
23. Olayode, I.O.; Tartibu, L.K.; Okwu, M.O. Prediction and modeling of traffic flow of human-driven vehicles at a signalized road intersection using artificial neural network model: A South African road transportation system scenario. *Transp. Eng.* **2021**, *6*, 100095. [[CrossRef](#)]
24. Qian, Y.; Zhang, X.; Fei, G.; Sun, Q.; Li, X.; Stallones, L.; Xiang, H. Forecasting deaths of road traffic injuries in China using an artificial neural network. *Traffic Inj. Prev.* **2020**, *21*, 407–412. [[CrossRef](#)]
25. Rahim, M.A.; Hassan, H.M. A deep learning based traffic crash severity prediction framework. *Accid. Anal. Prev.* **2021**, *154*, 106090. [[CrossRef](#)]
26. Fu, X.; Liu, J.; Jones, S.; Barnett, T.; Khattak, A. From the past to the future: Modeling the temporal instability of safety performance functions. *Accid. Anal. Prev.* **2022**, *167*, 106592. [[CrossRef](#)]
27. Afrin, T.; Yodo, N. A Long Short-Term Memory-based correlated traffic data prediction framework. *Knowl.-Based Syst.* **2022**, *237*, 107755. [[CrossRef](#)]
28. Slimani, N.; Slimani, I.; Amghar, M.; Sbiti, N. Road traffic forecasting using a real data set in Morocco. *Procedia Comput. Sci.* **2020**, *177*, 128–135. [[CrossRef](#)]
29. Naqvi, N.K.; Quddus, M.A.; Enoch, M.P. Do higher fuel prices help reduce road traffic accidents? *Accid. Anal. Prev.* **2020**, *135*, 105353. [[CrossRef](#)]
30. Katrakazas, C.; Michelaraki, E.; Sekadakis, M.; Ziakopoulos, A.; Kontaxi, A.; Yannis, G. Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. *J. Saf. Res.* **2021**, *78*, 189–202. [[CrossRef](#)]
31. Park, J.; Abdel-Aty, M.; Wang, J.H. Time series trends of the safety effects of pavement resurfacing. *Accid. Anal. Prev.* **2017**, *101*, 78–86. [[CrossRef](#)]
32. Vipin, N.; Rahul, T. Road traffic accident mortality analysis based on time of occurrence: Evidence from Kerala, India. *Clin. Epidemiol. Glob. Health* **2021**, *11*, 100745.
33. Roland, J.; Way, P.D.; Firat, C.; Doan, T.N.; Sartipi, M. Modeling and predicting vehicle accident occurrence in Chattanooga, Tennessee. *Accid. Anal. Prev.* **2021**, *149*, 105860. [[CrossRef](#)] [[PubMed](#)]
34. Shannon, D.; Fountas, G. Extending the Heston model to forecast motor vehicle collision rates. *Accid. Anal. Prev.* **2021**, *159*, 106250. [[CrossRef](#)]
35. Al-Hasani, G.; Khan, A.M.; Al-Reesi, H.; Al-Maniri, A. Diagnostic time series models for road traffic accidents data. *Int. J. Appl. Stat. Econom.* **2019**, *2*, 26.
36. Rashidi, M.H.; Keshavarz, S.; Pazari, P.; Safahieh, N.; Samimi, A. Modeling the accuracy of traffic crash prediction models. *IATSS Res.* **2022**, *in press*. [[CrossRef](#)]
37. Lunacek, M.; Williams, L.; Severino, J.; Ficenc, K.; Ugirumurera, J.; Eash, M.; Ge, Y.; Phillips, C. A data-driven operational model for traffic at the Dallas Fort Worth International Airport. *J. Air Transp. Manag.* **2021**, *94*, 102061. [[CrossRef](#)]
38. Sarmento, P.; Motta, M.; Scott, I.J.; Pinheiro, F.L.; de Castro Neto, M. Impact of COVID-19 lockdown measures on waste production behavior in Lisbon. *Waste Manag.* **2022**, *138*, 189–198. [[CrossRef](#)]

39. Melchior, C.; Zanini, R.R.; Guerra, R.R.; Rockenbach, D.A. Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. *Int. J. Forecast.* **2021**, *37*, 825–837. [CrossRef]
40. Chang, F.; Huang, H.; Chan, A.H.; Man, S.S.; Gong, Y.; Zhou, H. Capturing long-memory properties in road fatality rate series by an autoregressive fractionally integrated moving average model with generalized autoregressive conditional heteroscedasticity: A case study of Florida, the United States, 1975–2018. *J. Saf. Res.* **2022**, *in press*. [CrossRef]
41. Barcellos, R.; Bernardini, F.; Viterbo, J. Towards defining data interpretability in open data portals: Challenges and research opportunities. *Inf. Syst.* **2022**, *106*, 101961. [CrossRef]
42. Feng, Y.; Shah, C. Unifying telescope and microscope: A multi-lens framework with open data for modelling emerging events. *Inf. Processing Manag.* **2022**, *59*, 102811. [CrossRef]
43. De Souza, A.A.C.; d'Angelo, M.J.; Lima Filho, R.N. Effects of Predictors of Citizens' Attitudes and Intention to Use Open Government Data and Government 2.0. *Gov. Inf. Q.* **2022**, *39*, 101663. [CrossRef]
44. Gutierrez–Osorio, C.; Pedraza, C. Modern data sources and techniques for analysis and forecast of road accidents: A review. *J. Traffic Transp. Eng.* **2020**, *7*, 432–446. [CrossRef]
45. Veljković, N.; Bogdanović–Dinić, S.; Stoimenov, L. Benchmarking open government: An open data perspective. *Gov. Inf. Q.* **2014**, *31*, 278–290. [CrossRef]
46. Portal Otvorenih Podataka, Podaci o Saobraćajnim Nezgodama za Teritoriju Grada Beograda. Available online: <https://data.gov.rs/sr/datasets/podatsi-o-saobratshajnim-nezgodama-za-teritoriju-grada-beograda/> (accessed on 15 December 2021).
47. Portal Otvorenih Podataka. Podaci o Saobraćajnim Nezgodama po Policijskim Upravama i Opštinama. Available online: <https://data.gov.rs/sr/datasets/podatsi-o-saobratshajnim-nezgodama-po-politsijskim-upravama-i-opshtinama/> (accessed on 15 December 2021).
48. Pravno Informacioni System. Zakon o Elektronskoj Upravi. Službeni Glasnik 27. 6 April 2018. Available online: <http://www.pravno--informacioni--sistem.rs/SIGlasnikPortal/eli/rep/sgrs/skupstina/zakon/2018/27/4/reg> (accessed on 15 December 2021).
49. Sekadakis, M.; Katrakazas, C.; Michelaraki, E.; Kehagia, F.; Yannis, G. Analysis of the impact of COVID–19 on collisions, fatalities and injuries using time series forecasting: The case of Greece. *Accid. Anal. Prev.* **2021**, *162*, 106391. [CrossRef]
50. Box, G.E.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.
51. Sun, J.; Tang, M. The Programming Languages: Introduction of R. *Syst. Med. Integr. Qual. Comput. Approaches* **2021**, *1*, 1–8.
52. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; Available online: <http://www.stats.ox.ac.uk/pub/MASS4/> (accessed on 15 December 2021).
53. Trapletti, A.; Hornik, K. Tseries: Time Series Analysis and Computational Finance. R Package Version 0.10–50. Available online: <https://cran.r-project.org/web/packages/tseries/index.html> (accessed on 15 December 2021).
54. Hyndman, R.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O'Hara-Wild, M.; Petropoulos, F.; Razbash, S.; Wang, E.; Yasmeeen, F. Forecast: Forecasting Functions for Time Series and Linear Models. R Package Version 8.13. Available online: <https://pkg.robjhyndman.com/forecast/> (accessed on 15 December 2021).
55. Stoffer, D. Astats: Applied Statistical Time Series Analysis. R Package Version 1.12. Available online: <https://CRAN.R-project.org/package=astsa> (accessed on 15 December 2021).
56. Lütkepohl, H.; Xu, F. The role of the log transformation in forecasting economic variables. *Empir. Econ.* **2012**, *42*, 619–638. [CrossRef]
57. Lewis, C.D. *Industrial and Business Forecasting Models*; Butterworths: London, UK, 1982.
58. Yaffee, R.A.; McGee, M. *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*; Academic Press Inc.: Orlando, FL, USA, 2000.
59. Knežević, N.; Glišović, N.; Milenković, M.; Bojović, N. Prognoziranje prihoda od poštanskih usluga korišćenjem neuronskih mreža zasnovanih na metaheuristikama. In Proceedings of the XXXVI Simpozijum o Novim Tehnologijama u Poštanskom i Telekomunikacionom Saobraćaju—PosTel 2018, Beograd, Serbia, 4–5 December 2018; pp. 33–42.
60. Milenković, M.; Bojović, N. *Handbook of Research on Emerging Innovations in Rail Transportation Engineering*; Railway Demand Forecasting; Rai, B., Umesh, Eds.; IGI Global: Hershey, PA, USA, 2016; pp. 100–129. [CrossRef]
61. Towards Data Science. An Overview of Time Series Forecasting Models. Available online: <https://towardsdatascience.com/an-overview-of-time-series-forecasting-models-a2fa7a358fcb> (accessed on 23 March 2022).
62. García-Ferrer, A.; Bujosa, M.; de Juan, A.; Sánchez-Mangas, R. Revisiting the relationship between traffic accidents, real economic activity and other factors in Spain. *Accid. Anal. Prev.* **2020**, *144*, 105549. [CrossRef]
63. Ramírez, A.F.; Valencia, C. Spatiotemporal correlation study of traffic accidents with fatalities and injuries in Bogota (Colombia). *Accid. Anal. Prev.* **2021**, *149*, 105848. [CrossRef]
64. Comi, A.; Polimeni, A.; Balsamo, C. Road Accident Analysis with Data Mining Approach: Evidence from Rome. *Transp. Res. Procedia* **2022**, *62*, 798–805. [CrossRef]