DOI: https://doi.org/10.24297/jam.v23i.9656

A Stochastic Model to Predict Road Accidents in Albania

Denisa Kaçorri (Salillari)[1], Anita Caushi[2], Albina Basholli[3], Luela Prifti[4]

[1,2,3,4]Department of Mathematical Engineering, Polytechnic University of Tirana, ALBANIA

Abstract

The importance of predicting accident rates lies in the improvement of road infrastructure and the effective implementation of laws and traffic regulations. The statistical change of many phenomena over time is described by time series. This paper aims to forecast the number of individuals involved in road accidents in Albania by applying SARIMA model approach. This study used monthly number of individuals involved in road traffic accidents in Albania from 2016 to 2023. Using forecasting techniques for the number of traffic accidents can serve as a valuable strategy for achieving various objectives including the implementation of traffic safety campaigns, strategies, and action plans outlined in traffic safety initiatives. The model was found to be effective in capturing the underlying patterns and trends in the data, providing valuable insights for understanding and forecasting traffic accident occurrences in Albania.

Keywords: Time series, Forecast, SARIMA, Accidents, Albania.

Introduction

The statistical change of many phenomena over time is described by time series. They find a wide range of applications in various fields like physics, biology, telecommunication, trade, insurance, road traffic, industry, machine learning etc., (Chatfield, C., & Xing, H. 2019), (Montgomery, D. C., et al. 2024), (Box, G.E, et al. 2015). There are some applications of time series forecasting models in modelling road traffic and predicting number of deaths in road accidents (Deretić. N, et al., 2022), (Farida Merabet, Halim Zeghdoudi.,2020), (Theodor D. Popescu. 2020). (Getahun, K.A. 2021), and (Rabbani, M.B.A., et al. 2021). Statistical analysis of road accidents remains a vital component in the development of new road safety programs. By analysing the time series of road traffic accidents, we can identify trends, seasonal variations, and other patterns that are essential for informed decision-making. The alarming prevalence of road accidents worldwide, including in Albania, is deeply concerning. Each year, these accidents lead to numerous deaths and significant property damage. According to INSTAT (INSTAT, 2024), the primary cause of death after natural causes is accidents, which experienced a 17% increase in 2023 compared to 2022. The study of road accidents serves not only to gather information on accident statistics but also to forecast future accident rates. Kaçorri (Salillari), D., et al. (2023) presented a non-homogeneous Poisson process model with intensity function a polynomial function for annual data in road accidents in Albania. The model was built based on yearly data of the total number of accidents in Albania from 2014 to 2022. As conclusion the model in Kaçorri (Salillari), D., et al. (2023) predicted the number of individuals involved in road accidents to varies from 1621 to 1871 per year. This estimation signifies for an almost three years period. A fact is that the number of individuals involved in road accidents in Albania resulted 1731, (INSTAT, 2024). In this paper we apply the Box–Jenkins seasonal autoregressive integrated moving average model (SARIMA). Applying this model to the monthly road traffic accident data in Albania, spanning from 2016 to 2023, the study aims to provide a reliable forecast of the number of individuals involved in road accidents. When examining the data on the number of traffic accidents in Albania, it was found that there is a noticeable seasonal variation. This is why the SARIMA model was selected, as the ARIMA model is unable to account for both seasonality and trend in datasets.

Materials and Methods

The Autoregressive Integrated Moving Average (ARIMA) is one of the most used forecasting methods for time series data. SARIMA, also known as Seasonal ARIMA, represents a generalization of ARIMA model that facilitates the explicit modelling of the seasonal aspect within the data series. In this paper, the SARIMA model is used for short-term forecasts. A SARIMA model with non-seasonal terms of order (p, d, q) and seasonal terms of order (P, D, Q) is denoted a SARIMA(p, d, q) × (P, D, Q)s model and can be written

CC BY

$$\varphi(B)\Phi\left(B^s\right)(1 - B)^d\left(1 - B^s\right)^D Y_t = \theta(B)\Theta\left(B^s\right)X_t$$

where parameters ϕ and φ are the ordinary ARMA coefficients, θ and Θ are the seasonal ARMA coefficients, B is the backshift operator, whose effect on a time series Y_t can be summarized. θ, Θ denote polynomials in B^s of order P, Q, respectively, s is the length of the periodicity (seasonality) and X_t is a white noise sequence.

The main advantage of SARIMA is that the model is deterministic and computationally easy, (Box, G.E, et al. 2015), (Deretić. N, et al., 2022). Furthermore, the model effectively captures non-stationary behaviours within and across seasons through multiple parameterizations. However, limitations of SARIMA model include its ability to predict only a short period of time, and its constraint in extracting only linear relationships within the time series data, (Deretić. N, et al., 2022).

Results and Discussion

Data on the number of accidents in Albania was obtained from the Institute of Statistics INSTAT (INSTAT, 2024) The data presented in Table 1, provide information on the total number of individuals involved in road accidents that includes the number of injuries and fatalities within a specific timeframe from 2016 to 2023. This study used monthly road traffic accidents in Albania from 2016 to 2023.

The collected data was analysed and processed using the statistical program R.

Table 1: Number of individuals involved in road accidents in Albania

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2016	199	172	183	222	224	257	281	294	265	257	197	227
2017	158	159	213	190	240	245	272	292	214	223	235	170
2018	175	148	160	168	201	217	210	268	218	193	171	162
2019	153	154	176	147	143	187	228	230	152	143	144	187
2020	122	144	94	67	86	143	168	156	184	148	139	147
2021	165	103	150	145	151	161	175	200	141	179	156	134
2022	147	100	145	132	156	141	150	152	134	105	117	120
2023	119	123	118	94	136	132	165	180	122	130	212	200

Source: INSTAT (INSTAT, 2024)

CC BY

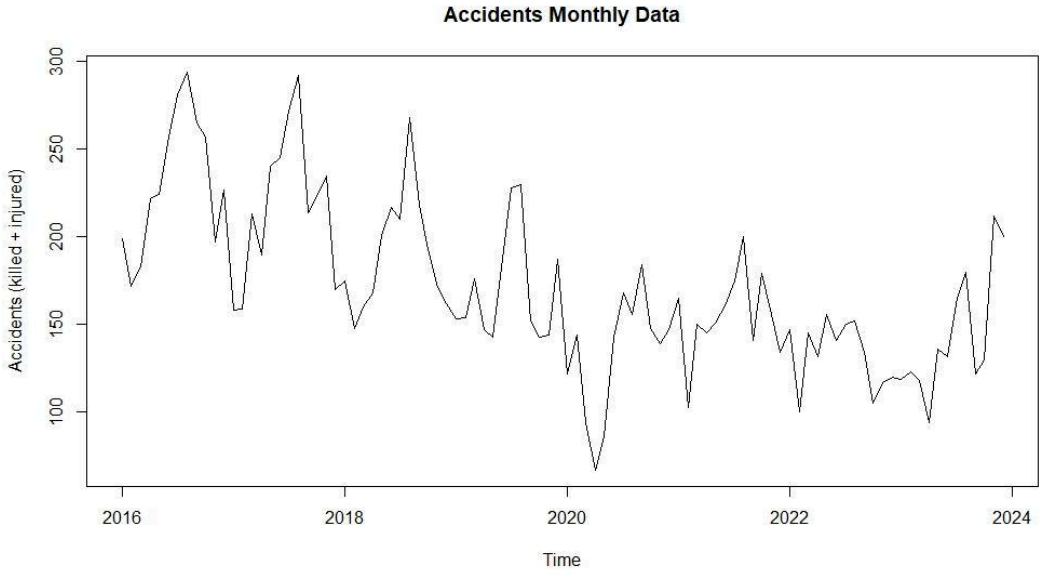

Figure 1: Monthly road traffic accident data in Albania, from 2016 to 2023.

From Figure 1, we observe a decrease in road accidents after 2019. The lowest values occur during the lockdown months due to the COVID-19 pandemic.

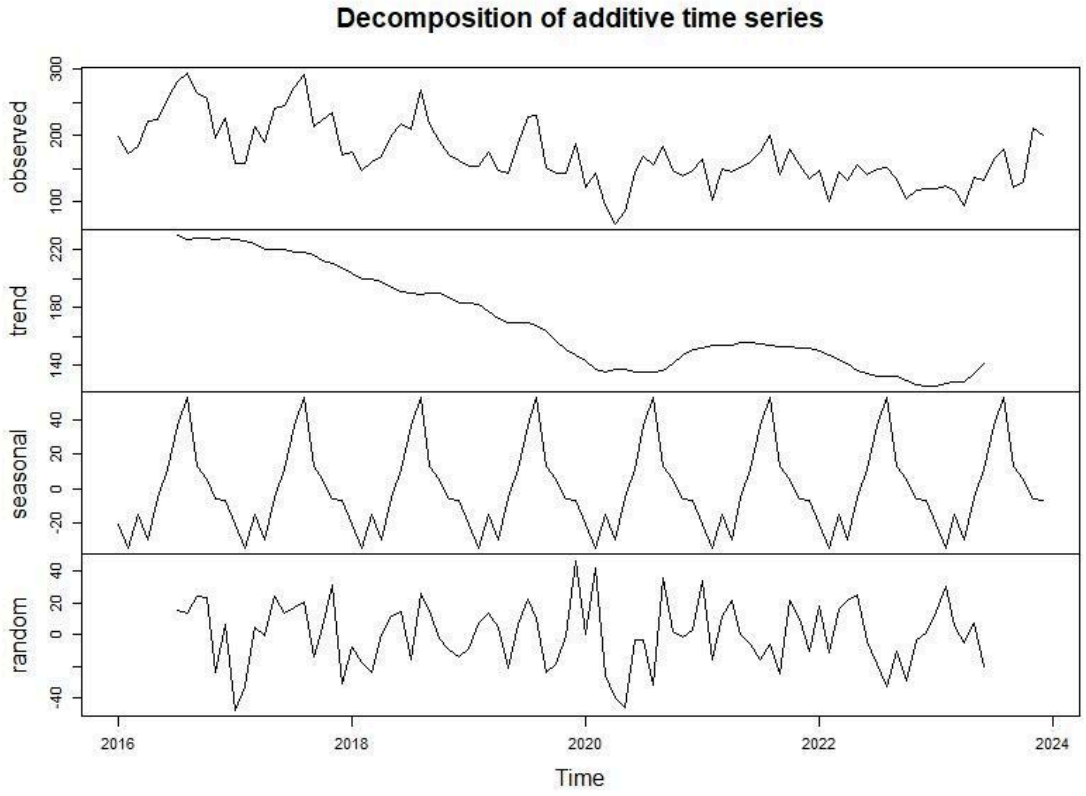

Figure 2: The decomposition of time series of road traffic accident data in Albania.

In the data presented in Figure 2, it is evident that there is a clear seasonal pattern, with peak indicators occurring from June to September, particularly in August, and with lower indicators in February and April. These fluctuations are likely related to the influx of tourism during the summer holidays.

CC BY

Albania is located in the Balkans stretches nearly 300 miles along the Mediterranean coast on the Adriatic and Ionian Seas. The country's coastline length is about 481 km (299 mi) composed of both sandy and rocky beaches, bays, caves, cliffs, capes and lagoons which make the main attractions of tourism in Albania.

Using the `arima()` function from stats package in R, we fitted an Autoregressive Integrated Moving Average model to the data. The model selected is an ARIMA with drift (SARIMA), which suggests that the series exhibits a seasonal pattern with a period of 12 months, along with an autoregressive component and a moving average component.

The summary of the fitted model is as follows:

> summary(arima_model)

Series: accidents

ARIMA(1,0,0)(0,1,1)[12] with drift

Coefficients:

	ar1	sma1	drift
	0.4966	-0.8815	-1.1059
s.e.	0.1037	0.2813	0.1903

sigma^2 = 740.4: log likelihood = -403.46

AIC=814.91 AICc=815.42 BIC=824.64

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-1.865251	24.99473	18.2725	-2.853039	12.33605	0.5640903	-0.06139571

From the results we can write the SARIMA model as:

$$y(t) = 1.1059 + 0.4966 * y(t-1) - 0.8815 * \Delta y(t-1) + \varepsilon(t) - \theta_1 * \Delta\varepsilon(t-1)$$

From the model the value of the autoregressive parameter (ar1) is 0.4966, which indicates a moderate positive correlation between the current and past values of the number of the accidents. The seasonal moving average parameter is -0.8815, suggesting a strong negative correlation between the current and past seasonal components. The drift coefficient of -1.1059 indicates a decreasing trend in the number of accidents over the study period. The model has a mean absolute percentage error (MAPE) of 12.33% and can be seen as an indicator that the prognosis is acceptably accurate. The time series model developed in this study can be used to forecast future accident trends in Albania and support the development of targeted interventions to improve road safety. In figure 3 is presented the time series with the forecasting values of the number of accidents during 2024. The lowest values are predicted during April 2024 with a 95% interval estimation]28.63606; 153.3754[, and the highest values are estimated during August 2024 with a 95% interval]97.65623; 222.6235[. The forecast values from SARIMA model for year 2024 are presented in Table 2. From the forecast, it is expected that the number of accidents in Albania will decrease in 2024.

Future research should focus on incorporating additional other variables, such as weather conditions, economic factors, and infrastructure changes, to further enhance the predictive capabilities of the model.

CC BY

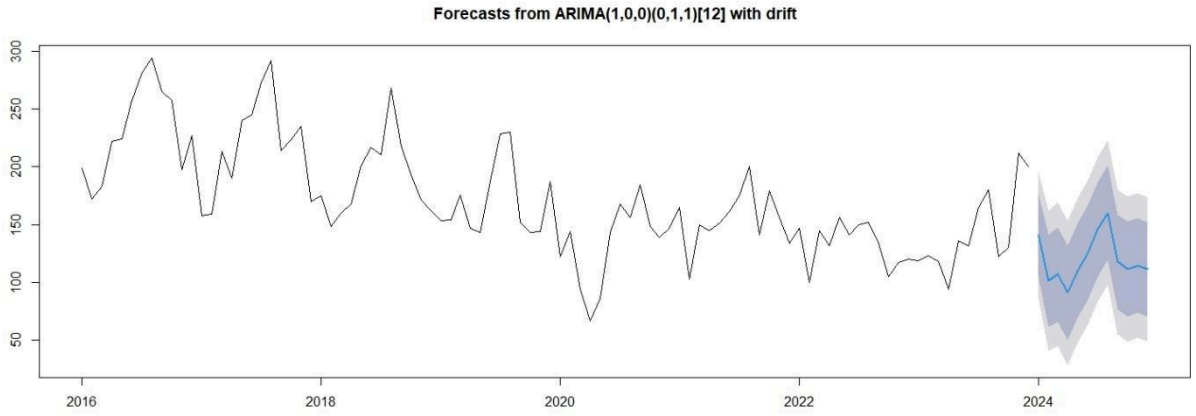

Figure 3: Forecasts from the SARIMA model.

Table 2: Forcast values of the Sarima model and their 95% confidence intervals

Month	Forecast value	Interval estimation 95%	
		L	H
Jan 2024	140.78844	86.47637	195.1005
Feb 2024	101.31616	40.74445	161.8879
Mar 2024	106.95026	44.93209	168.9684
Apr 2024	91.00575	28.63606	153.3754
May 2024	110.31894	47.86288	172.7750
Jun 2024	125.52576	63.04843	188.0031
Jul 2024	145.75857	83.27604	208.2411
Aug 2024	160.13988	97.65623	222.6235
Sep 2024	117.69963	55.21637	180.1829
Oct 2024	111.54840	49.06794	174.0289
Nov 2024	114.47939	52.01058	176.9482
Dec 2024	111.12762	48.70614	173.5491

Conclusions

In this paper, we presented a time series analysis approach to model and forecast road accidents in Albania using monthly data from 2016 to 2023. The SARIMA model demonstrated a good fit to the data, capturing the underlying patterns, trends and seasonality in road accidents. The model provides valuable insights for understanding and forecasting traffic accident occurrences in Albania. From the forecast, it is expected that the number of accidents in Albania will decrease in 2024.

CC BY

References

Box, G.E.; Jenkins, G.M.; Reinsel, G.C. (2015). Time Series Analysis: Forecasting and Control; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 5 ed.

Chatfield, C., & Xing, H. (2019). The Analysis of Time Series: An Introduction with R (7th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781351259446

Deretić, N., Stanimirović, D., Awadh, M. A., Vujanović, N., & Djukić, A. (2022). SARIMA Modelling Approach for Forecasting of Traffic Accidents. Sustainability, 14, 4403. https://doi.org/10.3390/su14084403

Farida Merabet, Halim Zeghdoudi.. (2020). On Modelling seasonal ARIMA series: Comparison, Application, and Forecast (Number of Injured in Road Accidents in Northeast Algeria). WSEAS Transactions on Systems and Control. 2020;15:235-246. 10.37394/23203.2020.15.25

Getahun, K.A. (2021). Time series modeling of road traffic accidents in Amhara Region. J Big Data 8, 102. https://doi.org/10.1186/s40537-021-00493-z

INSTAT (2024), https://www.instat.gov.al/al/temat/industria-tregtia-dhe-sh%C3%ABrbimet/transporti-aksidentet-dhe-karakteristikat-e-mjetet-rrugore/#tab2M.

Kaçorri (Salillari), D., Basholli, A. and Prifti, L. (2023) "Non-homogeneous Poisson Process with polynomial function rate to predict road accidents: A case study in Albania", AS-Proceedings, 1(2), pp. 446–450. doi: 10.59287/as-proceedings.194

Montgomery, D. C., Jennings, C. L., Kulahci, M. (2024). Introduction to Time Series Analysis and Forecasting. United States: Wiley.

Rabbani, M.B.A., Musarat, M.A., Alaloul, W.S. et al. (2021). A Comparison Between Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ES) Based on Time Series Model for Forecasting Road Accidents. Arab J Sci Eng 46, 11113–11138. https://doi.org/10.1007/s13369-021-05650-3

Theodor D. Popescu. (2020). Time Series Analysis for Assessing and Forecasting of Road Traffic Accidents - Case Studies. WSEAS Transactions on Mathematics. 19:177-185. 10.37394/23206.2020.19.17

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

CC BY