Time series analysis of Luanda road accidents, deaths and injureds

Manuel Alberto[1,2], Dulce Gomes[2,3], Patrícia A. Filipe[2,3]

[1] Instituto Superior Politécnico Internacional de Angola, Luanda, Angola
[2] Centro de Investigação em Matemática e Aplicações, IIFA, Universidade de Évora, Portugal
[3] Departamento de Matemática, ECT, Universidade de Évora, Portugal

E-mail for correspondence: d11837@alunos.uevora.pt

Abstract: In this work, time series models are applied to explain and forecast the rate of traffic accidents, deaths and injureds in Luanda, Angola. Monthly Luanda data from 2002 to 2015 are used to fit models and to make predictions. Road accidents in Angola are currently one of the major causes of death in the country. Particularly Luanda, the capital, is the province that shows the highest rate in terms of accidents, deaths and injureds. However, in recent years there has been a decrease in the accidents rate, with average growth rates of -6.73%, 0.19% and -2.54% for accidents, deaths and injureds respectively. We have used classic Seasonal ARIMA models (SARIMA) in two different approaches, the first one treat all observations the same way. The second approach identifies outliers, taking into account its magnitude and estimates SARIMA models for the series excluding the significant outliers. A Seasonal-Trend decomposition based on a locally-weighted regression smoothing (Loess) approach was also applied. The SARIMA models that take into account the extreme values revealed to fit and predict better than the pure SARIMA models time series of traffic accident data.

Keywords: Road accidents, Outliers, Time series, Seasonal ARIMA models, Decomposition.

1 Introduction

The overall statistical data of Direcção Nacional de Viação e Trânsito (DNVT) and Gabinete de Estudos, Informação e Análise (GEIA) show that in recent years, the road accidents rate in Angola has been decreasing in Luanda. However, current levels remain worrying given that the province has the highest rate at the country level in the time horizon from 2002 to

This paper was published as a part of the proceedings of the 34th International Workshop on Statistical Modelling (IWSM), University of Minho, Portugal, 7-12 July 2019. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

2015 (DNVT (2018)). During the armed conflict in Angola, there was a massive exodus of the population of the remaining provinces of Angola to Luanda, capital of the country. Even with the end of the conflict, due to lack of conditions in the other provinces, the population remains concentrated in the capital, with approximately 7 million inhabitants, in an area of 18 817,77 km^2 (information from Census2014). This concentration causes the anarchic growth of the city, without order in the traffic, which causes huge disturbances in the circulation, and consequently the high number of road accidents and their consequences. Urban growth has become incompatible with the road network, and the lack of an adequate public transport network, contributing in part for the increase in the number of accidents (DNVT (2018)). Recently, the Conselho Nacional de Viação e Ordenamento do Território indicates that amongst others, the causes of road accidents in Luanda, are factors such as the lack of illumination of the public highway, the drunkenness of motorists, the poor state of the roads and even the vehicles, etc.

In early work (Alberto *et al.*, (2018)) we have presented some results on modelling the rates of road accidents, deaths and injureds in Angola through classic seasonal ARIMA models (SARIMA), exploring and comparing two different approaches. One in which all observations are treated the same way, and the other that identifies outliers, taking into account its magnitude, and adjust SARIMA models excluding the significant outliers. The best models were identified and used to fit and predict road accidents, deaths and injured in Angola. Since, Luanda is the capital and the most problematic area, in this work we intend to develop a more detailed analysis for this province adapting and applying similar methodology as used before. We aim to be able to contribute with valuable information that can be useful to create measures to reduce the number of cases of road accidents. All analyzes were performed using statistical software `R` version 3.4.3.

2 Data and Methodology

Data used in this study was collected from Direcção Nacional de Viação e Trânsito (DNVT) and Gabinete de Estudos, Informação e Análise (GEIA) databases. These data refer to accidents, deaths and injured recorded from 2002 to 2015. The study was made using different time series methodologies. To study the trend of the road accidents, deaths and injured in Angola, we have used the STL (Seasonal-Trend decomposition by Loess) procedure (Cleveland (1990)), based on a locally-weighted regression smoothing (Loess) approach, that decompose data into trend, seasonal and irregular components. The STL filtering method is much more general than the classical one. It allows, for example, multiple seasonal and cyclical components and linear and non-linear trends due to the non-parametric nature of this method. The addictive decomposition was assumed. The span of loess window for seasonal extraction was set as 7, hence allowing the seasonal component to change over the years (Cleveland *et al.* (1990)). To model the monthly rates of road accidents, deaths and injured, we have used SARIMA models and the Box and Jenkins methodology to describe the autocorrelation in the observed data (see, *e.g.*, Brockwell and Davis (1991)).

In our data several upper outliers were observed in an early stage exploratory analysis. It is well known that in the time series context outliers and structural changes may arise, it is important that the models can incorporate these characteristics. The effects of outliers are, however, often omitted due to lack of knowledge of the methods to detect them and to accommodate them to the time series modeling process. Different procedures for detection and correction of outliers have been proposed in the literature (see, *e.g.*, Chen and Liu (1993) and Alberto *et al.*(2018)). An equation of the model in the presence of outliers for the series can be written as:

$$Y_t = \sum_{j=1}^{m} w_j \xi_j(B) I_t(t_j) + X_t$$

where $\{X_t\}$ is the SARIMA process, m is the number of detected outliers, w_j is the magnitude of the jth outlier in time t_j, I_t is an indicator function that assumes value 1 if $t = t_j$ and value 0 otherwise, $\xi_j(B)$ determines the dynamic of the outlier in time $t = t_j$ identifying the type of outlier (see, *e.g.*, Alberto *et al.* (2018)). A test set with the monthly rates of road accidents, deaths and injured in 2015 was used to evaluate the models forecasts, through accuracy measures like: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).

3 Main Results and Conclusions

We used the `tso` function from the `tsoutlier`R package, to identify the outliers, estimate its magnitude and fit a SARIMA model to the series without the identified outliers. The best fitted models were SARIMA$(1,1,1)$ $(2,0,2)_{12}$, SARIMA$(0,1,1)(0,0,1)_{12}$ and SARIMA$(2,0,0)(1,0,0)_{12}$ for road accidents, deaths and injured in Luanda from 2002 until 2015. Using the classical SARIMA approach, the parsimonious models that best fitted the monthly rates was SARIMA$(0,1,2)(0,1,3)_{12}$, SARIMA$(2,0,0)(1,0,0)_{12}$ and SARIMA$(1,1,1)(1,0,2)_{12}$ for road accidents, deaths and injured, respectively. Both estimated models fit well enough, capturing the variability of rates over the period considered. However, as expected, fitting through classical models do not capture so well the abrupt changes of the observed values.

With respect to forecasting, we have left out the last observed year, as a test set to validate forecasts. For both approaches, the models were readjusted for the period 2002-2014 and predicted the monthly rates for 2015. The models obtained for both approaches were similar to those estimated for the complete period. To evaluate the forecast, accuracy measures previously mentioned were computed. The following table shows these results.

TABLE 1. RMSE, MAE and MAPE values for fitting and forecasts of the monthly rates of road accidents, deaths and injured in Luanda.

		Classical approach		Outliers approach	
		Fitting	Forecast	Fitting	Forecast
Accidents	RMSE	2.29	2.14	1.46	1.47
	MAE	1.39	1.56	1.14	1.28
	MAPE	21.77	25.45	21.60	54.83
Deaths	RMSE	0.37	0.37	0.33	0.30
	MAE	0.28	0.28	0.27	0.23
	MAPE	19.02	18.45	18.91	24.19
Injured	RMSE	1.85	1.92	1.11	0.93
	MAE	1.22	1.26	0.85	0.78
	MAPE	30.78	30.32	22.52	39.73

The Seasonal ARIMA models that take into account the extreme values revealed to fit and predict better than the classical Seasonal ARIMA models time series of traffic accident data.

Acknowledgments: The authors belong to research centre CIMA - Universidade de Évora supported by FCT (UID/MAT/04674/2019).

References

Alberto, M., Gomes, D. and Filipe, P.A. (2018). A time series approach to road accidents in Angola from 2002 to 2015. *International Journal of Applied Mathematics and Statistics*, **57**, 63 – 73.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Method*. New York: Springer-Verlag.

Chen, C. and Liu, L. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of American Statistical Association*, **88**, 284 – 297.

Cleveland, R. and Cleveland, W. and Mcrae, J. and Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, **6**, 3 – 73.

DNVT (2018). *Relatório sobre a Sinistralidade Rodoviária em Angola*. Direcção Nacional de Viação de Trânsito, Angola.