Application of the Hybrid Forecasting Models to Road Traffic Accidents in Algeria

Fatih Chellai[1] | *Ferhat Abbas University, Setif, Algeria*

Received 3.11.2021 (revisions received 3.12.2021, 12.2.2022), **Accepted** (reviewed) 15.3.2022, **Published** 17.6.2022

Abstract

Road traffic accidents are a growing public health concern. In this study, we focused on analyzing and forecasting the monthly number of accidents, number of injuries, and number of deaths in Algeria over the period (2015–2020). For this purpose, hybrid forecasting models based on equal weights and in-sample errors were fitted, and we compared them with the seasonal autoregressive moving average (SARIMA) models. The three models retained for forecasting until 2022 are all hybrid models, one based on equal weight and two models based on in-sample errors (using the RMSE indicator). Furthermore, the hybrid models outperformed the SARIMA models for short (6 months), medium (12 months), and long horizon (24 months). The forecasting results showed that we expect an increase in the number of accidents, the number of deaths, and the number of injuries over the next 12 months. Policymakers must enhance strategies for prevention and road safety, especially in rural areas, where the highest rate of fatalities is recorded.

Keywords	***DOI***	***JEL code***
Road traffic accidents, hybrid forecasting models, seasonal time series analysis	*https://doi.org/10.54694/stat.2021.37*	*C22, C32, C53*

INTRODUCTION

Road traffic accidents are a real public health issue, adding that, its negative effects go beyond the health dimension to the social and economic dimensions Racioppi et al. (2004). According to the World Health Organization (WHO), road traffic accidents (RTA) cause 1.3 million fatalities with more than 5 million people injured during 2020, see WHO (2021). The other dark side is the social and economic consequences of road traffic accidents which include degradation of the quality of life and psychiatric impacts for the victim and its family (Mayou et al., 1993), loss of productivity, the cost of the legal system, and medical costs (Ansani et al., 2020; Bardal and Jørgensen, 2017; Chen et al., 2019).

Specifically, Algeria is in the top ranking of the most affected developing countries owing to road traffic accidents. Statistics show an increase of 42.6% in the number of RTAs in the first five months of 2021 compared with 2020. The same tendency was recorded for the number of injuries and the number

[1] Department of Based Education, Ferhat Abbas University, Setif 1, El Bez Campus, 19000, Algeria. E-mail: fatih.chellai@univ-setif.dz, phone : (+213)0663526184.

of fatalities, which increased by 40.8% and 21.6%, respectively, compared to the first five months in 2020. On the other side, the number of vehicles in Algeria was estimated at more than 6.4 million in 2018; in the same year 255 538 new vehicles have been registered. However, according to the data delivered by the National Office of Statistics ONS Algeria (2021) there were 6.1 million cars in 2017, representing an annual increase of 4.1%. The challenge of reducing the number of accidents has led many researchers to provide solutions to determine the effective factors in accident occurrence and to present comprehensive safety programs and strategies. The main aim of this study is to demonstrate the optimality and accuracy of hybrid models in forecasting the trajectories of the number of RTAs, mortalities, and injuries in Algeria.

To the best of the author's knowledge, few studies have applied hybrid methods to predict the patterns of road traffic accidents. This study is a new step in providing high-quality statistics in terms of reliability and regularity over a relatively long period, which can help decision-makers monitor and evaluate the efficiency of prevention strategies. The key objectives are (*i*) to explore the patterns of road accidents in Algeria by considering the spatial dimension, and (*ii*) to introduce hybrid models for forecasting the trajectories of the number of accidents, the number of injuries, and deaths.

The remainder of this article is organized as follows: the next section discusses the most frequently used models in forecasting. Second section presents the main features of the hybrid forecasting models. The third section describes the RTA data for Algeria. The fourth section presents the main results of modeling and forecasting. The last two sections discuss, conclude and summarize the findings of the study.

1 OVERVIEW OF FORECASTING METHODS

With the development of computer and simulation techniques, several statistical models (linear and non-linear) have emerged and have been applied in the field of modeling and forecasting of time series data, the most important of which is the Box-Jenkins method Box and Jenkins (1970), which is a useful linear model that has proven its efficiency and importance in the field of forecasting Ihueze and Onwurah (2018). In a general context, Hyndman and Athanasopoulos (2018) provide a good reference for the best practices and forecasting principles. We also mention the book of the time-series analysis by Hamilton (2020). On the other hand, with the development of machine learning and big data, non-linear models such as artificial neural networks (ANN) have emerge and have also been applied with great acceleration over the past years, principally by Delen et al. (2006) and Rezaie et al. (2011). However, in practice, we face several challenges in choosing the optimal model for data and prediction, and we use information criterions as well as accuracy measures for forecasting. In light of this choice, the approach of merging these candidate models has emerged to reach a single prediction result, which we call the combined forecasting method, see Granger and Ramanathan (1984), and Armstrong (2001). The application of this approach is abundant in different fields; Zhang (2003) applied a hybrid method by combining ARIMA and artificial neural networks (ANNs) and concluded that such a combination can improve the forecasting accuracy better than the single original method. In the same axis Wang et al. (2013) combined ARIMA and the ANNs and tested to forecast different datasets. Abdollahi (2020) applied a hybrid model to forecast the dynamics of oil prices. We also mention the study carried out by Rezaie et al. (2011), who applied artificial neural networks to predict the severity of road accidents, and revealed that several factors, such as human factors and weather factors could increase the crash severity in urban highways. Using the ARIMA and ARIMAX models, Ihueze and Onwurah (2018) attempted to predict the trajectories of road accidents in Anambra State (Nigeria). They concluded that transfer function models (ARIMAX) are preferred over ARIMA models.

Yusuf et al. (2015) used fuzzy logic to develop a hybrid approach for forecasting enrolment and car road accidents. The findings of the study showed that the presented method performs better the forecasting results comparing to other existing methods. Barba et al. (2014) presented a combination of ARIMA models and autoregressive neural networks (ANNs) to improve the forecasting of traffic accidents. Following a two-step strategy of combination, they revealed that an ARIMA-HSVD (Hankel matrix) performed better in the forecasting results compared with other combinations. In recent study Sangare et al. (2021) developed a new combination framework with a Gaussian mixture model (GMM) and a support vector Classifier (SVC) to forecast urban traffic, they revealed that the new hybrid approach performed better than the road accident baseline statistical models.

2 HYBRID FORECASTING MODELS

Recently, the approach of forecast combining has been widely applied in different fields of research, the main idea of which is to combine the forecasts from different techniques such as ARIMA models, ETS (Error, Trend, Seasonal), and ANN,... the process of combination is based on the weights of each technique to the final forecast output; in practice, we can select the weights through in-sample errors, which was first introduced by Bates and Granger (1969), and the other way to select the weights is by cross-validation. The theoretical background of our study is based on the work of Bates and Granger)1969), which was the original study on this topic. In the same context Yang (2004) revealed that empirical research advocates that hybrid models usually improve forecasting accuracy over the original approaches.

For our application, we used five forecasting methods: (1) the ARIMA model; (2) theta model, which was developed by Assimakopoulos and Nikolopoulos (2000); (3) neural network models Hill et al. (1996) based on feed-forward and sing hidden layers and lagged inputs; (4) the exponential smoothing state-space model (ETS) Hyndman et al. (2002); and (5) the TBATS model (exponential smoothing state-space model with Box-Cox transformation, ARMA errors, trends, and seasonal components), De Livera et al. (2011).

In this study, we are interested in forecasting the number of road accidents, number of deaths, and number of injuries. We define ψ as a forecasting method that provides us the forecasts of $\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_k}$ at different horizons up to k, we can simply estimate the average risk $R(\psi; k)$ to measure the accuracy of the ψ method. To achieve this objective, we have a class of forecasting Ω that contains several statistical approaches $\Omega = \{\psi_1, \psi_2, ..., \psi_m\}$, where: m may be finite or infinite. After forecasting, a probable departure e_t of the predicted values $\widehat{y_t}$ from the real valuescan occur, and the optimal approach is that gives us the minimal discrepancy e_t.

In the context of the hybrid forecasting models, any forecasting technique ψ is called a combined forecasting procedure if the forecasts outcomes $\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_k}$ constitute a measurable function on the real values $y_1, y_2, ..., y_k$ and the values of $\hat{y}_{i,j}$ where $1 \le i \le$ k and $1 \le j \le m$.

In a general form, the model that provides the combination of these forecasting methods can be defined as follows:

$$\widehat{y_t} = \sum_{i=1}^{n} w_i \widehat{y_i}(t),$$

where w_i are the weights designed for each forecasting method, generally we have $\sum_i w_i = 1$, but this is not always the case, see for more details Lean et al. (2005).

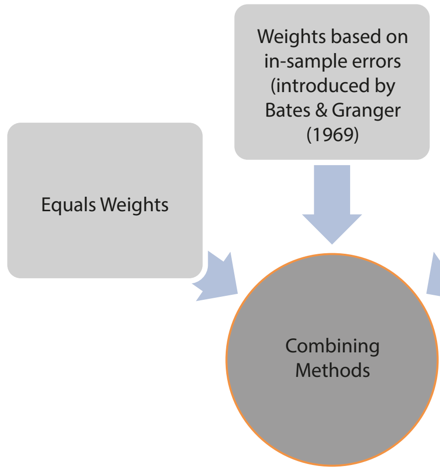
The process of combination can be applied with three approaches:

(1) Equal weights (i.e.) $w_i = \frac{1}{n}$; which is a standard and robust method as depicted by Lean et al. (2005).

(2)Weights based on in-sample errors; the main idea of this method is to resolve a system of equations (generally quadratic programming) to find at the end the optimal weights,

$$\begin{cases} Min\left(w_i f(e_t)\right) \\ \sum_{i=1}^{m} w_i = 1 \quad , w_i \ge 0 \; .i = 1,2,\ldots,m \quad .t = 1,2,\ldots,T. \end{cases}$$

Figure 1 Hybrid forecasting models



Source: Own construction

The idea of this method was firstly introduced by Bates and Granger (1969), and we have several options to choose the function $f(e_t)$ as an indication, the R package we work on provides three functions (or errors measures): MAE: Mean Absolute Error, MASE: Mean Absolute Scaled Error and RMSE: Root Mean Square Error.

(3) Cross-validated weights; cross-validation of time series data with user-supplied models and forecasting functions is also supported to evaluate the model accuracy.

The comparison (and model evaluation) between different combined forecasting classes and between the SARIMA models was conducted using the following accuracy measures: the root mean squared errors, $RMSE=\sqrt{\frac{\sum_t (y_t-\hat{y}_t)^2}{n}}=\sqrt{\frac{\sum_t (e_t)^2}{n}}$. The mean absolute errors, $MAE=\frac{\sum_t |y_t-\hat{y}_t|}{n}=\frac{\sum_t |e_t|}{n}$. The maximum of the absolute percentage error, $MAPE=Max\left(\left|\frac{y_t-\hat{y}_t}{y_t}\right|\right)\times 100$. The mean percentage errors,

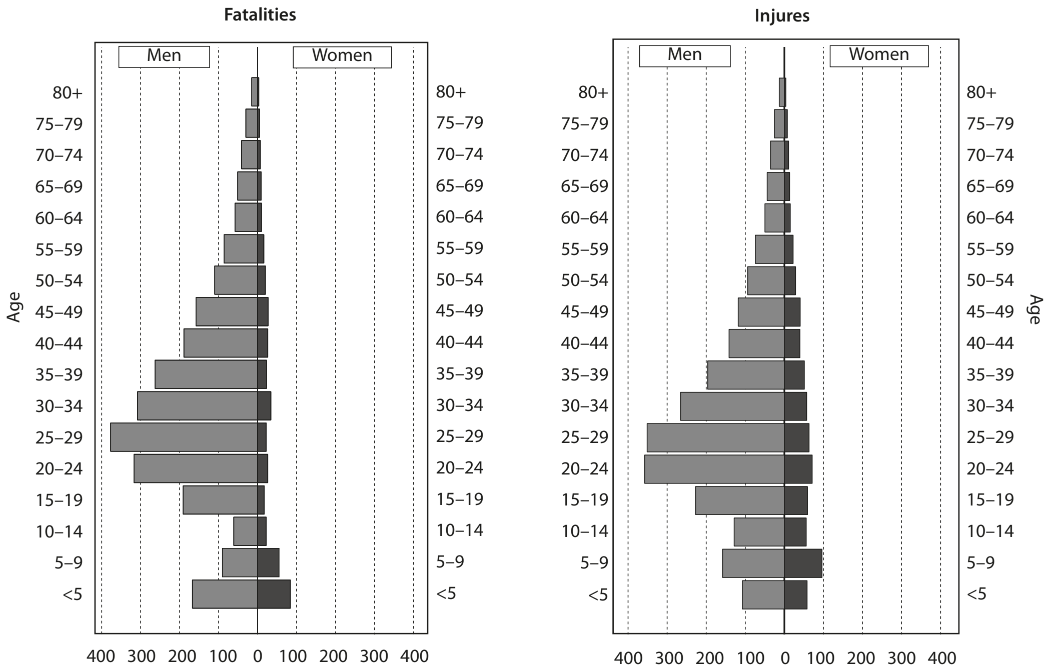
$$MPE=\frac{100\%}{n}\sum_{t=1}^{n}\left(\frac{y_t-\hat{y}_t}{y_t}\right).$$

3 AN OVERVIEW OF ROAD TRAFFIC ACCIDENTIN ALGERIA

To analyze the trajectories of road traffic accidents in Algeria, we used data that are provided by the DNSR (Délégation Nationale à la Sécurité Routière) in 2020. The entire dataset was completed after obtaining a license and an official request was submitted to the administrative body of the Ministry of the Interior. In practice, however, the process of collecting information on traffic accidents is undertaken by the Civil Protection, the National Gendarmerie (in rural areas), and the police (in urban areas), after this step, the DNSR mission is to prepare, clean, and organize these data and elaborate periodical reports about the road traffic accidents in the country.

As a global fact, road traffic injuries are among the foremost factors of mortality in the general population. Specifically, children and young adults aged 5–29 years were most affected by these accidents. This is true for Algeria, as the pyramids in Figure 2 showed the distribution of deaths and injuries according to sex and age. We can see that the most affected age-category is 25–29 years for both sexes. Figures in 2020, and as a cumulative frequency, 43.6 % of the deaths and 55.9% of injuries were among persons aged under 29 years. Regardless of the age category, gender statistics show that males are more likely to be involved in road traffic crashes than females. Approximately three-quarters (73.1%) of all road traffic deaths occur among young men under the age of 25 years, who are almost three times as likely to be killed in a road traffic crash as young women.

Figure 2 Distribution of traffic accident victims by age and sex in 2020



Source: Author's computations based on data provided by the DNSR (2020)

According to statistics provided by the Office National des Statistiques (ONS Algeria), the total number of under-five child mortality was 22 240 in 2019 (for both sexes). The data delivered by the DNSR showed that the number of fatalities among children under five years of age was 25. Consequently, 1.1% of the total deaths among children under five years of age were caused by traffic accidents. Furthermore, the mortality rates were much higher for boys than girls. As conjectural figures, and based on the recent statistics provided by the DNSR in June 2021, we found a clear increase either in the number of accidents and the number of death and injuries, this is shown in Table 1.

Table 1 Evolution of the number of crashes, number of injuries, and number of deaths between 2020 and 2021

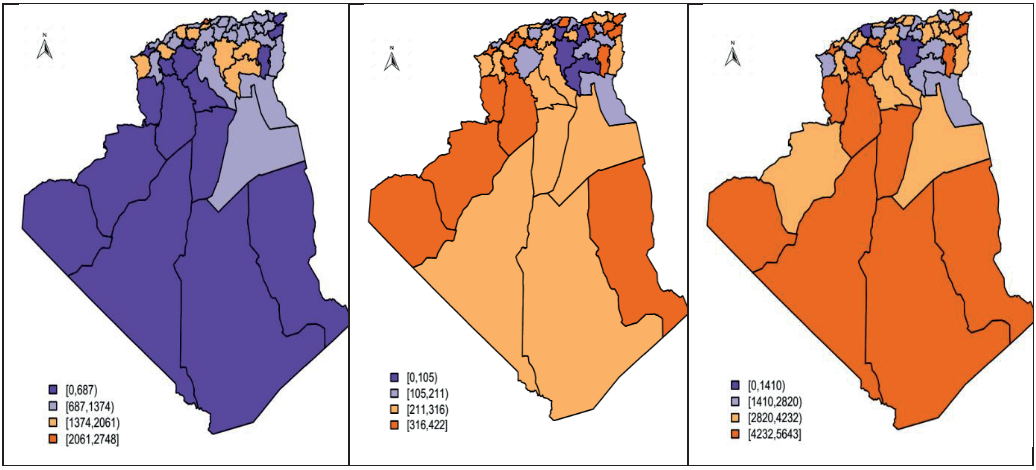
Year	Accidents	Injuries	Deaths
2020	7 216	9 708	1 065
2021	10 292	13 664	1 295

Source: Author's computations based on data provided by DNSR (2020, 2021)

In the first five months of 2020, the number of accidents was 7 216, but in the same period in 2021, the number of accidents was 10 292, with an increase of 42.6 % compared to 2020, the same tendency was recorded for the number of injuries and the number of fatalities, respectively, with an increase of 40.7 %

and 21.6% compared to the first five months in 2020. This high variation between the two periods can be explained by the strategies of the containment due to Covid-19 taking by the government at the beginning of 2020, which has been (after July 2020) removed (partially) by allowing traffic between and among states.

Figure 3 Spatial distribution of the number of crashes, number of injuries, and number of deaths in Algeria



Source: Author's computations based on data provided by the DNSR (2020)

There was a significant difference in the number of accidents, the number of fatalities, and the number of injuries and across the 48 states of Algeria, as shown in the three maps in Figure 2. Furthermore, we estimated the road fatality rate (RFT), and heterogeneity still existed among the states. This spatial analysis can be used as an indicator of comparison among regions and allows policymakers to develop suitable strategies for each region to improve road safety in the country. If detailed statistics for the 48 states of Algeria are available, we believe that spatial modeling approaches, such as geographically weighted regression (GWR), could provide more insights into the spatial differentiation of road traffic accidents in these states, and a recent study conducted by Wachnicka et al. (2021) showed the reliability of this statistical method to identify regional differences in road traffic accidents in Europe.

At another decomposition level, statistics showed that the number of accidents in urban areas was 15 211, representing 66.2% of the total number of accidents. This was twice the number of accidents in rural areas. Compared to 2019, the number increased by 5.1% in urban areas and decreased by 16.3% in rural areas. However, the number of deaths due to traffic accidents is mainly in rural areas. In 2019, the total number of deaths due to road accidents was 2 599, that was 79.4% of the total number, four times higher than in urban areas. The rural traffic network is characterized by ease of traffic, which allows high speeds and a low level of surveillance, representing the greatest challenge in terms of security.

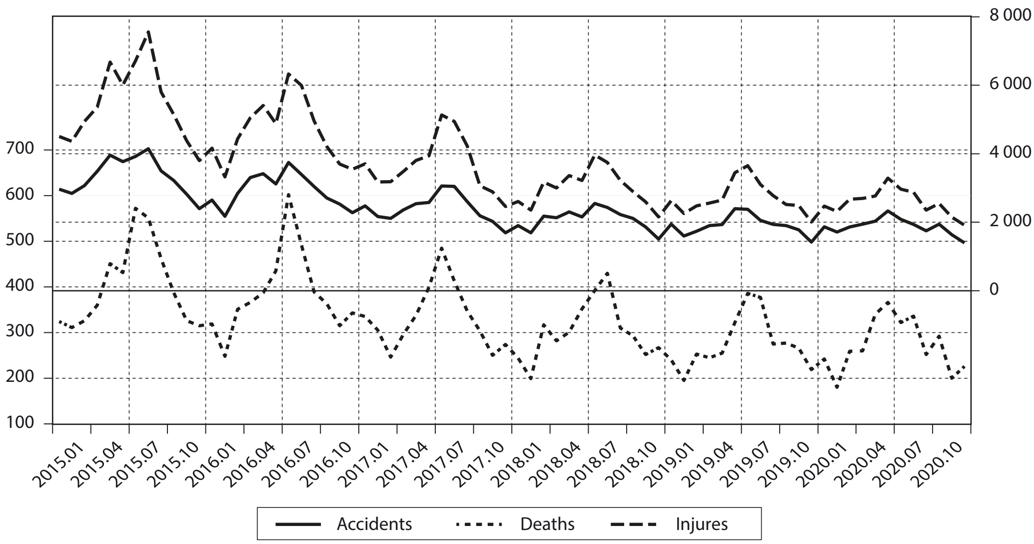
4 RESULTS

4.1 Stationary time series analysis

The plots in Figure 3 describe the trajectories of the number of accidents, deaths, and injuries in Algeria over the period (2015–2020) as the monthly frequency. Graphically, the three variables exhibited the same pattern over the study period, with a slightly decreasing trend over the period in which a seasonal component in the time series was clearly observed. The descriptive statistics showed that the average number (per month) of accidents was 2 426, of deaths, and 328 and 3 682 injuries per month, respectively. Based on the coefficient of variation (CV), we found the same level of dispersion for the number of accidents and number of deaths (27.3% and 26.5 %, respectively). In contrast, the highest dispersion was recorded for the number of injuries, with a coefficient of variation of 35.1%. We noted a slight decrease in the number of road traffic accidents in the last year (2020), as well as in the number of injuries and deaths. This decrease may be due to lockdown strategies caused by the Covid-19 pandemic, as health authorities in Algeria have taken several safety measures to reduce transportation by 2020, which has witnessed a large spread of the virus.

An analysis of the outliers was conducted in order to test the validity of the hypothesis of the effect of the lockdown on the reduction in the number of accidents. Specifically, we aim to investigate the presence of level shifts, transient changes, and innovation outliers in the time series over the ten last months (March 2020 to December 2020). For this objective, we used the "tsoutliers" (v0.6-8; Javier, 2019) R package, which was developed on the approach of Chen and Liu (1993). Except for a transient change in the "Deaths" time series that was identified in 2020, the results revealed no presence of a significant switching in level for the three time series during the last year (i.e. in 2020). In contrast, before 2020, the test results showed the presence of a seasonal-level shift (SLS) in the number of "Accidents" time series. At the same time, transient and level changes were observed in the "Injuries" time series during 2016 and 2017, more details are in the Appendix 3.

Figure 4 Time series plots of the evolution of the number of accidents, number of deaths, and number of injuries in Algeria over the period (2015–2020)



Source: Author's computations based on data provided by the DNSR (2020)

For the normality assumption, we conducted statistical tests on the stationary time series, see the Appendix 2. The kernel densities in the left and right axis borders of the plots confirmed the non-normality distribution of the stationary deaths time series and the normality of the injuries and accidents variables; these was also tested by the test of normality of Jarque and Bera. As detailed information about the shape of the data distribution, the Fisher skewness parameters showed that the number of deaths time series followed an asymmetric distribution (skewed left), the kurtosis coefficients indicated that the dispersion of the extremes values is higher in the time series of "number of deaths" compared to the other variables (number of accidents and number of injuries), a detailed reference in measuring skewness is conducted by Doane and Seward (2011).

Since the data doesn't exhibit the trend, and this is for the three variables (Accidents, Injuries, and Deaths), we select to test the unit root hypothesis for "intercept" only, for the optimal lag selection in testing we follow the automatic option based on the Akaike Information Criterion (AIC), Akaike (1974). The critical values of the test were based on simulations, and all statistical programs provided critical values at different levels of significance, according to the sample size. In the literature, we find several seasonal unit root tests, but the most commonly used is the Hylleberg, Engle, Granger, and Yoo (HEGY) test; see Hylleberg et al. (1990). For application, Ronderos (2019) provided a simple and comprehensive procedure using the Eviews program.

Table 2 Seasonal Unit Root Test for the three time-series using the HEGY method

All seasonal frequencies	Tabulated test statistics at different significance level and different sample size			Calculated test statistics		
	1%	5%	10%	Accidents	Deaths	Injuries
n = 40	28.09	7.38	3.43	2.6123	2.7475	3.9065
n = 60	28.137	7.36	3.49			
n = 53*	28.12	**7.37**	3.47			

Note: n = 53* included observations after adjustment, which are obtained using linear interpolation. For the frequencies, typically, we worked on 0-frequency, $2\pi/4$, $6\pi/4$ and π.
Source: Author's calculation

The null hypothesis of this test states that a unit root exists at a specified frequency periodicity. To test this hypothesis, we select the option "all frequencies", and we worked on adjusted sample size n = 53 which is close to the lower and upper values of the simulation, in our case, the lower is n = 40 and the upper is n = 60. Thus, because the critical values of the test statistics (2.61, 2.74, and 3.90) were smaller than the critical value (7.37) at the 5% significance level, the null hypothesis is accepted. The seasonality in the statistical series is expressed by the third quarter corresponding to major holiday trips, and the fourth quarter to social re-entry and the onset of bad weather as the most dangerous periods in terms of road traffic. In general, seasonality in road traffic accidents is partly due to traffic trends and partially due to weather conditions.

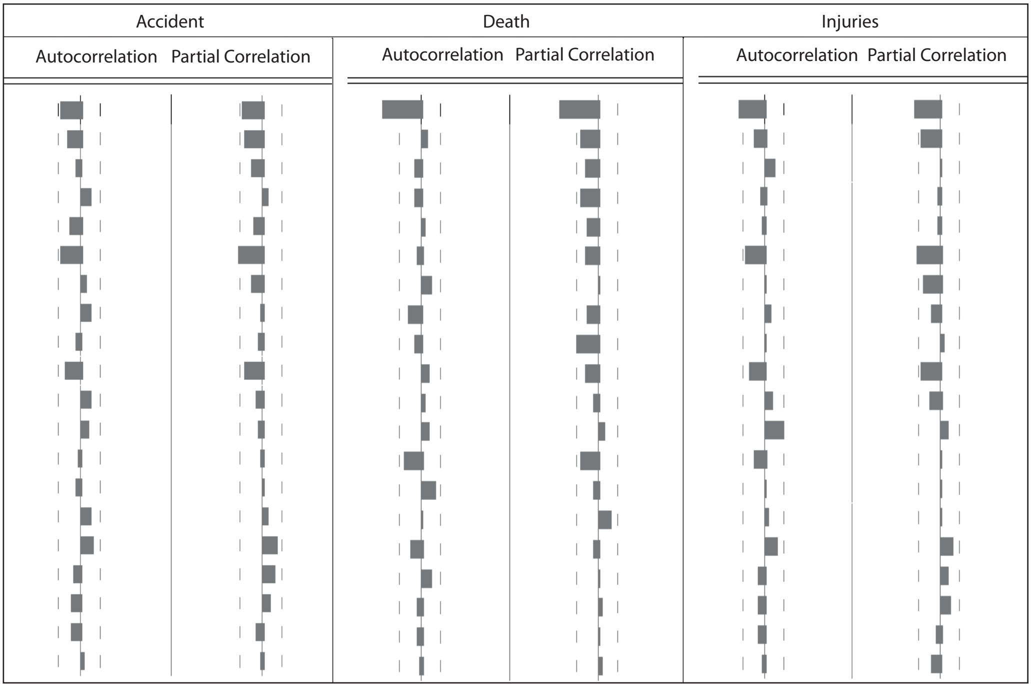
Table 3 The Augmented Dicky-Fuller Unit-Root test

Series	*p. value* at level			*p. value* at first (trend) difference		
	None	Intercept	Intercept & Trend	None	Intercept	Intercept & Trend
Accidents	0.3421	0.0684	0.2244	0.0000	0.0001	0.0004
Fatalities	1	0.1361	0.9971	0.3138	0.0154	0.0203
Injuryies	1	0.9986	0.0171	0.5541	0.0016	0.0056

Source: Author's calculation

As shown in the last two columns of Table 3, the stationarity assumption of the three time series is confirmed by the computed p-values of the Augmented Dicky-Fuller test, which are all lower than the (0.05) significance level. Furthermore, this stationarity was confirmed by the autocorrelation (ACF) and partial autocorrelation (PACF) functions shown in Figure 4, which behave as stationary processes. More precisely, the plots of these functions show the absence of autocorrelation(moving average component) and partial autocorrelation (autoregressive component) in the series of accidents (Figure 5(a)) .The two components (moving average and autoregressive) are statistically significant for the series of deaths and injuries and (b) and (c), respectively.

Figure 5 Autocorrelation and partial auto-correlations functions of the first difference of time series



Source: Author's construction

4.2 Forecasting results and model comparison

After data preparation and stationarity analysis, the main thing remaining to achieve is model identification, selecting the optimal model, and forecasting the trends of the three variables. First, for the model combination (and weight method), we worked on two methods: equal weight ($w_i = \frac{1}{n}$), and in-sample errors. As indicated in the method section, we worked on five principal methods of forecasting; this selection is justified by two reasons: these methods are the most used in forecasting of time series, and also they are all available in the "forecastHybrid" (v.5.0.19; Shaub and Ellis, 2020) R package which helps researchers to usethis combining approach in other studies.

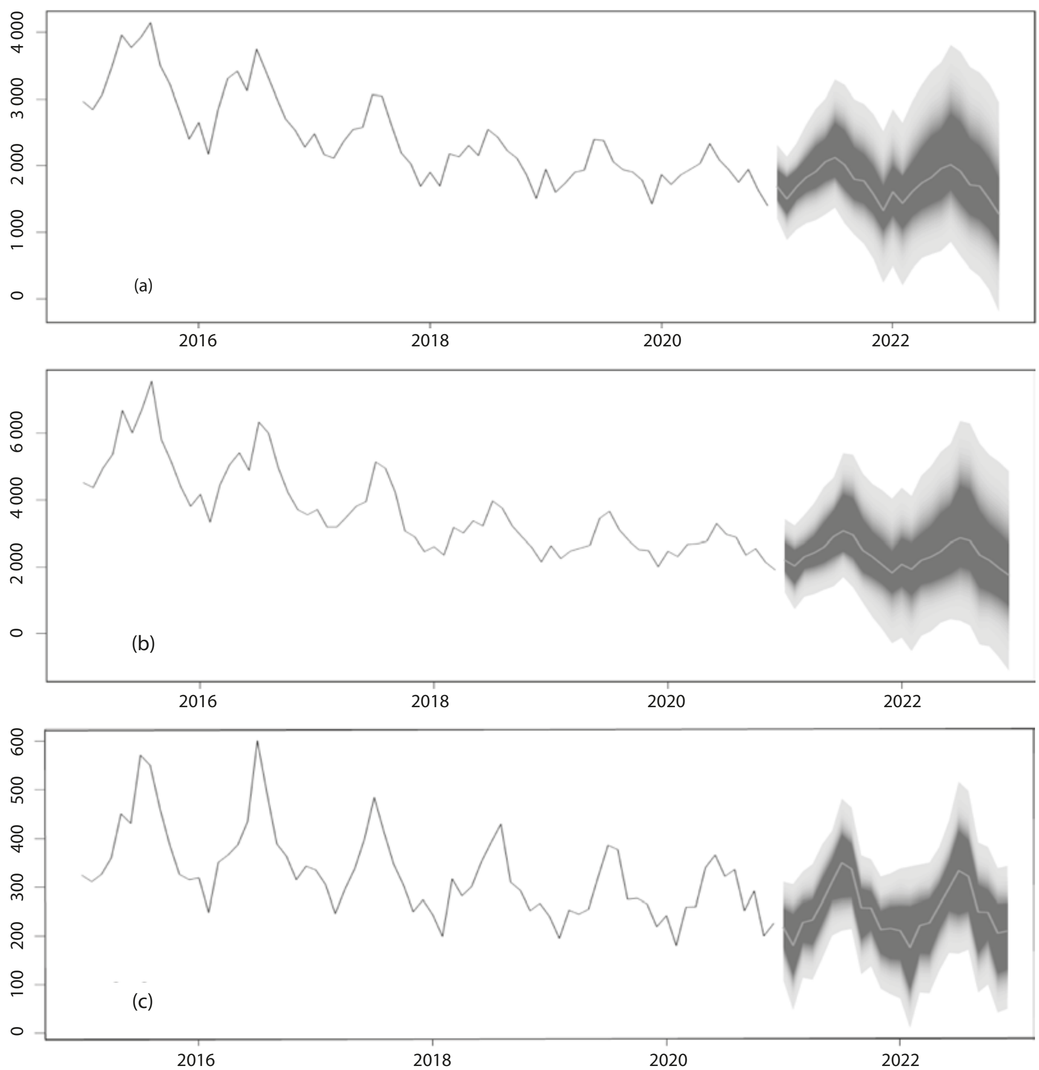
Table 4 Accuracy of the Hybrid models comparing with SARIMA models

Variables	Models	ME	RMSE	MAE	MPE	MAPE	ACF1
Accidents	Hybrid-M	–15.344	159.366	127.216	–1.2002	5.721	0.049
	SARIMA	19.963	179.506	129.932	1.161	5.814	0.407
Injuries	Hybrid-M	–44.197	268.981	209.042	–2.067	6.419	0.068
	SARIMA	15.121	329.972	229.096	0.972	6.973	0.383
Deaths	Hybrid-M	0.4302	33.4053	24.8803	0.014	7.969	0.614
	SARIMA	–5.3106	30.4623	23.2241	–3.031	7.921	0.006

Note: ME – Mean Error, **RMSE** – Root Mean Squared Error, **MAE** – Mean Absolute Error, **MPE** – Mean Percentage Error, **MAPE** – Mean Absolute Percentage Error, **ACF1** – Autocorrelation of errors at *lag 1*.
Source: Author's computation based on R program

Table 4 presents the accuracy measures of the optimal hybrid-models and the SARIMA models. The performances indicators (ME, RMSE, MAE, MPE, MAPE, and ACF1) showed that the hybrid models outperformed the SARIMA models for the three variables(accidents, injuries, and deaths). For the ARIMA models, the optimal ones were: a SARIMA$(1,1,1)(1,1,0)_{12}$ for the accident variable, and a SARIMA$(0,1,1)(2,1,0)_{12}$ for the injuries variable and a SARIMA$(1,0,0)(1,1,0)_{12}$ with drift for deaths variable, detailed characteristics of the selected SARIMA model are in the Appendix 1. The predictions estimated by the hybrid-models and SARIMA models were validated using the dataset of the last 12 months (in 2020). The models retained for forecasting are all hybrid-models; one based on equal weight for the accident variable and the two rest models based on in-sample errors (of RMSE indicator).

Figure 6 Hybrid-models forecasts for the number of accidents (a), number of Injuries (b), and number of deaths (c)



Note: The dashed blue surfaces correspond to the upper and lower bounds of confidence intervals of prediction at the α = 0.1, and the dashed gray surfaces et the α = 0.05 significance levels of predictions.
Source: Author's plot using R program

As can be seen in Figure 5, we expect an increase in the number of accidents and the number of injuries and deaths, and we expect that the number of accidents in August 2021 to be (on average) 2011 accidents, 2 945 injuries, and 335 deaths. The forecasts for the coming year (2022) follow nearly the same trajectories for the three variables.

5 DISCUSSION

In this study, a descriptive and predictive analysis was carried out on road traffic accidents in Algeria over the period (2015–2020), where the hybrid forecasting models were estimated and compared with the Box-Jenkins models. The findings revealed that the combined methods outperformed the SARIMA models, and we expect an increase in the number of accidents, number of deaths, and number of injuries over the next 12 months. Detailed statistics and estimation results have been presented, but the challenge is how to transform these figures into strategies.

This study presents, for the first time, the application of hybrid models to forecast the trajectories of road traffic accidents in Algeria. Our findings suggest the optimality of the hybrid models over the Box–Jenkins model. Compared with previous studies, this finding is broadly consistent with the study by Barba et al. (2014), which revealed the performance of combining Hankel matrix (HSVD)-ARIMA models with ARIMA models in forecasting traffic accidents in Chile. Similar results have been reported by Yusuf et al. (2015). Recently, Sangare et al. (2021) stated that the approach of the combination forecasting method was more accurate than baseline statistical methods in forecasting urban traffic accidents.

It was found that the number of deaths due to road traffic accidents in rural areas was four times higher than that in urban areas. This result is in good agreement with those of previous studies. For example, Cabrera-Arnau et al. (2020) explored road accident data from England and Wales, and reported that fatal crashes were more likely in rural areas than in urban areas. This pattern was demonstrated by Darma et al. (2017) in Malaysia, who revealed that the number of traffic fatalities in rural zones (66% of total deaths) was higher than that in urban zones. Accordingly, by exploring national surveillance data in China, Wang et al. (2019) showed that rural areas have higher road traffic mortality rates than urban areas do.

The results showed that men (regardless of age) were more likely to be involved in RTAs than were women; this finding is consistent with previous studies on this topic. The same finding was reported by Razi-Ardakani et al. (2018), who confirmed that men had a higher risk of road traffic accidents and attempted to analyze the factors behind sex differences in traffic accident severity. By analyzing data on road traffic accidents in Ecuador, Algora-Buenafé et al. (2017) indicated that 81.1% of fatal traffic accidents corresponded to men and 18% to women. Similarly, Wang et al. (2019) revealed that men in China had higher road accident mortality rates than women.

In terms of outlier analysis, there was a significant transient change in the total number of deaths after the lockdown due to the Covid-19 pandemic; However, this result was not conclusive in the case of Algeria. By contrast, recent studies in other countries have demonstrated the effects of pandemics on road safety. For example, Katrakazas et al. (2020) reported that the number of road accidents in Greece was reduced by 41% during the lockdown. In the same issue, Saladié et al. (2020) stated that the daily number of accidents was reduced by 74.3% during the period of lockdown in Tarragona province, Spain.

CONCLUSION

Summing up the results, it can be concluded that this study has shown the past and future dynamics of road traffic accidents in Algeria in terms of the number of accidents, injuries, and deaths. The optimality of the hybrid models over the Box-Jenkins model for forecasting RTAs is demonstrated. However, further studies are required to determine the optimal number of forecasting methods for the combined process. Several other questions remain to be addressed in order to better understand the pattern of road traffic in Algeria. Specifically, future studies using regression models should be useful for estimating the effects of vehicle characteristics, road conditions, driver characteristics, and weather conditions on the dynamics of road traffic accidents in Algeria.

ACKNOWLEDGEMENT
This research received no specific grant from any funding agency, commercial or not-for-profit sectors. The author would like to thank the responsible of the Délégation Nationale à la Sécurité Routière (DNSR) to provide us the access to the dataset and other information on road Accidents. We would like to thank the two anonymous reviewers for helpful comments and improvements.

References

ABDOLLAHI, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied Energy*, 267: 115035.
AKAIKE, H. (1974), A new look at the statistical model identification [online]. *IEEE Transactions on Automatic Control*, 19(6): 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.
ALGORA-BUENAFÉ, A. F., SUASNAVAS-BERMÚDEZ, P. R., MERINO-SALAZAR, P., GÓMEZ-GARCÍA, A. R. (2017). Epidemiological study of fatal road traffic accidents in Ecuador. *Australasian Medical Journal*, 10(3): 238.
ARMSTRONG, J. S. (2001). Combining forecasts. In: *Principles of forecasting*, Boston, MA: Springer: 417–439.
ASSIMAKOPOULOS, V., NIKOLOPOULOS, K. (2000). The theta model: a decomposition approach to forecasting. *International journal of forecasting*, 16(4): 521–530.
BARBA, L., RODRÍGUEZ, N., MONTT, C. (2014). Smoothing strategies combined with ARIMA and neural networks to improve the forecasting of traffic accidents. *The Scientific World Journal*.
BARDAL, K. G., JØRGENSEN, F. (2017). Valuing the risk and social costs of road traffic accidents – Seasonal variation and the significance of delay costs. *Transport Policy*, 57: 10–19.
BATES, J. M., GRANGER, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4): 451–468.
BOX, G., JENKINS, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
CABRERA-ARNAU, C., PRIETO CURIEL, R., BISHOP, S. R. (2020). Uncovering the behaviour of road accidents in urban areas. *Royal Society open science*, 7(4): 191739.
CHEN, C., LIU, L.-M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series [online]. *Journal of the American Statistical Association*, 88(421): 284–297. <https://doi.org/10.2307/2290724>.
CHEN, S., KUHN, M., PRETTNER, K., BLOOM, D. E. (2019). The global macroeconomic burden of road injuries: Estimates and projections for 166 countries. *The Lancet Planetary Health*, 3(9): e390–e398.
DARMA, Y., KARIM, M. R., ABDULLAH, S. (2017). An analysis of Malaysia road traffic death distribution by road environment. *Sādhanā*, 42(9): 1605–1615.
DE LIVERA, A. M., HYNDMAN, R. J., SNYDER, R. D. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496): 1513–1527.
DOANE, D. P., SEWARD, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of Statistics Education*, 19(2).
GRANGER, C. W., RAMANATHAN, R. (1984). Improved methods of combining forecasts. *Journal of forecasting*, 3(2): 197–204.
HAMILTON, J. D. (2020). *Time series analysis*. Princeton university press.
HILL, T., O'CONNOR, M., REMUS, W. (1996). Neural network models for time series forecasts. *Management Science*, 42(7): 1082–1092.
HYLLEBERG, S., ENGLE, R. F., GRANGER, C. W., YOO, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1–2): 215–238.
HYNDMAN, R. J., KOEHLER, A. B., SNYDER, R. D., GROSE, S. (2002). A state space framework forautomatic forecasting using exponential smoothing methods. *International J. Forecasting*, 18(3): 439–454.
HYNDMAN, R. J., ATHANASOPOULOS, G. (2018). *Forecasting: principles and practice*. OTexts.
IHUEZE, C. C., ONWURAH, U. O. (2018). Road traffic accidents prediction modelling: an analysis of Anambra State, Nigeria. *Accident Analysis & Prevention*, 112: 21–29.
KATRAKAZAS, C., MICHELARAKI, E., SEKADAKIS, M., YANNIS, G. (2020). A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. *Transportation research interdisciplinary perspectives*, 7: 100186.

LEAN, Y., SHOUYANG, W. A. N. G., LAI, K. K., NAKAMORI, Y. (2005). Time series forecasting with multiple candidate models: selecting or combining? *Journal of Systems Science and Complexity*, 18(1): 1–18.
LÓPEZ DE LACALLE, J. (2019). *Tsoutliers: Detection of Outliers in Time Series* [online]. R package version 0.6-8. <https://CRAN.R-project.org/package=tsoutliers>.
MAYOU, R., BRYANT, B., DUTHIE, R. (1993). Psychiatric consequences of road traffic accidents. *British Medical Journal*, 307(6905): 647–651.
ONS. (2021). *Parc. Automobile* [online]. Office National des Statistiques, Algeria. <https://www.ons.dz/IMG/pdf/e.nat31_12_2018.pdf>.
RAZI-ARDAKANI, H., ARIANNEZHAD, A., KERMANSHAH, M. (2018). A Study of Sex Differences on Road Crash Severity. *Proceedings of the 3rd International Conference on Civil, Structural and Transportation Engineering (ICCSTE'18)*.
REZAIE MOGHADDAM, F., AFANDIZADEH, S., ZIYADI, M. (2011). Prediction of accident severity using artificial neural networks. *International Journal of Civil Engineering*, 9(1): 41–48.
RONDEROS, N. (2019). *Seasonal Unit Root Tests* [online]. <http://blog.eviews.com/2019/04/seasonal-unit-root-tests.html>.
SALADIÉ, Ò., BUSTAMANTE, E., GUTIÉRREZ, A. (2020). COVID-19 lockdown and reduction of traffic accidents in Tarragona province, Spain. *Transportation research interdisciplinary perspectives*, 8: 100218.
SANGARE, M., GUPTA, S., BOUZEFRANE, S., BANERJEE, S., MUHLETHALER, P. (2021). Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. *Expert Systems with Applications*, 167: 113855.
SHAUB, D., ELLIS, P. (2020). *Forecast Hybrid: Convenient Functions for Ensemble Time Series Forecasts* [online]. R package version 5.0.19. <https://CRAN.R-project.org/package=forecastHybrid>.
WACHNICKA, J., PALIKOWSKA, K., KUSTRA, W., KIEC, M. (2021). Spatial differentiation of road safety in Europe based on NUTS-2 regions. *Accident Analysis & Prevention*, 150: 105849.
WANG, L., ZOU, H., SU, J., LI, L., CHAUDHRY, S. (2013). An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioral Science*, 30(3): 244–259.
WANG, L., NING, P., YIN, P., CHENG, P., SCHWEBEL, D. C., LIU, J., HU, G. et al. (2019). Road traffic mortality in China: Analysis of national surveillance data from 2006 to 2016. *The Lancet Public Health*, 4(5): e245–e255.
WHO. (2021, June). *Road traffic injuries* [online]. Geneva: World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.
YUSUF, S. M., MU'AZU, M. B., AKINSANMI, O. (2015). A Novel Hybrid Fuzzy Time Series Approach with Applications to Enrollments and Car Road Accidents. *International Journal of Computer Applications*, 129(2): 37–44.
YANG, Y. (2004). Combining forecasting procedures: Some theoretical results. *Econometric Theory*, 20(1): 176–222.
ZHANG, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50: 159–175.

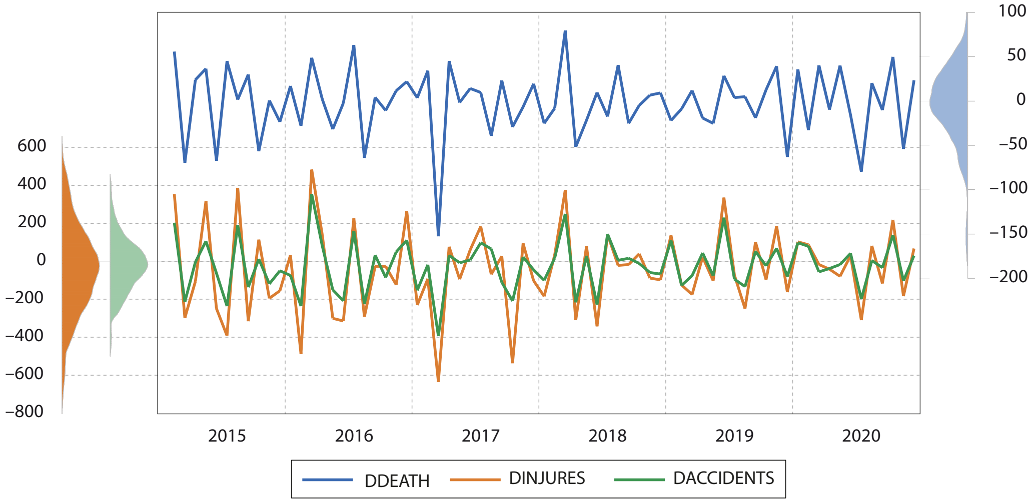
APPENDICES

Appendix 1 Characteristics of SARIMA models

Model	Variable	AIC	AICc	BIC	Variance	Log-Likelihood
ARIMA$(1,1,1)(1,1,0)_{12}$	Accidents	802.65	803.39	810.96	41429	–397.33
ARIMA$(0,1,1)(2,1,0)_{12}$	Injuries	876.03	876.77	884.34	139 991	–434.02
ARIMA$(1,0,0)(1,1,0)_{12}$ with drift	Deaths	613.98	614.71	622.36	1 410	–302.99

Note: SARIMA model is defined as: *ARIMA* $\underbrace{(p, d, q)}_{\text{(Non-seasonal part of the model)}} \underbrace{(P, D, Q)_m}_{\text{(Non-seasonal part of the model)}}$, where p: Trend Autoregressive component, d: Difference order, q: Moving Average component. The seasonal part of the model is designed as P: Seasonal Autoregressive component, D: Seasonal Difference order, Q: Seasonal Moving Average component, m: the frequency of the time series; here $m = 12$ which means monthly data and it can exhibits an annual seasonal cycle.
Source: Own construction

Appendix 2 Plot of stationary time series with kernel densities in the axis borders



Source: Own construction

Appendix 3 Summary of outliers' analysis

Variable	*id*	Type(*)	Time	Coef.	T-stat
Accident	1	AO	2016:02:00	–318.9	–3.394
	2	AO	2016:07:00	356.9	3.700
	3	TC	2017:03:00	–460.2	–3.639
	4	IO	2019:06:00	616.4	3.579
Injuries	1	SLS	2016:08:00	–972.5	–3.654
	2	TC	2017:03:00	–949.6	–3.969
	3	LS	2017:10:00	–779.1	–3.274
Deaths	1	LS	2017:03:00	–92.84	–8.440
	2	TC	2020:07:00	–106.34	–3.298

Note: (*) "AO" additive outliers, "LS" level shifts, "TC" temporary changes, "IO" innovative outliers and "SLS" seasonal level shifts.
Source: Own construction