

Estimating Traffic Accidents in Turkey Using Differential Evolution Algorithm

Ali Payıdar Akgüngör, Ersin Korkmaz

Kırıkkale University, Engineering Faculty
Department of Civil Engineering, 71451 Yahsihan, Kırıkkale, Turkey
e-mail: aakgungor@gmail.com, ersin_korkmaz1@hotmail.com

Abstract

Estimating traffic accidents play a vital role to apply road safety procedures. This study proposes Differential Evolution Algorithm (DEA) models to estimate the number of accidents in Turkey. In the model development, population (P) and the number of vehicles (N) are selected as model parameters. Three model forms, linear, exponential and semi-quadratic models, are developed using DEA with the data covering from 2000 to 2014. Developed models are statistically compared to select the best fit model. The results of the DE models show that the linear model form is suitable to estimate the number of accidents. The statistics of this form is better than other forms in terms of performance criteria which are the Mean Absolute Percentage Errors (MAPE) and the Root Mean Square Errors (RMSE). To investigate the performance of linear DE model for future estimations, a ten-year period from 2015 to 2024 is considered. The results obtained from future estimations reveal the suitability of DE method for road safety applications.

Key words: Traffic accident models, Differential Evolution Algorithm, Road Safety, Turkey

1. Introduction

Deaths and injuries caused by traffic accidents are serious traffic safety problems for not only developing countries but also developed countries. According to the statistics of World Health Organization (WHO), worldwide 1.3 million people die and as many as 20-50 million people are also injured annually because of traffic accidents. The same statistics expressed that 90 % of deaths resulted from traffic accidents occur in low and middle income countries, even though they own less than 50 % of the motor vehicles in the world [1].

Traffic accidents, injuries and fatalities are also a serious public health problem in Turkey. According to the statistics obtained from Turkey Statistical Institute (TSI), approximately 1.2 million traffic accidents occurred in the country, and more than 3.5 thousand people lost their lives and about 285 thousand people were injured in these accidents in 2014 [2]. These figures are much higher than those of European countries. The main reason behind this figure is due to the fact that freight and passenger traffic are concentrate on highway transportation system. More specifically, today 95% of passenger and 90% of freight transportation are managed on highways in Turkey. Additionally, growing population, developing economy and increasing

the number of vehicles in the country are other reasons on traffic accidents. Considering the general situation of the country during the period of 2000-2014, the population has gone up from 67.4 to 77.7 million while the number of vehicles has increased from 8.3 to 18.8 million. During the same period, the number of accidents have raised from 0.5 to 1.2 million.

Accident prediction models are employed by transportation planners, decision makers and engineers to develop new strategies for road safety. One of the oldest models is accident prediction model developed by Smeed in 1949. The author investigated the relationship among the population, the number of fatalities and vehicles by using the data gathering from various 20 countries [3]. However, Andreassen criticized Smeed's model because he was used only one-year data and each country has different social and economic conditions. He proposed a different exponential fatality model for each country [4]. Mekky [5] investigated the relationship between the ratio the number of vehicles to the number of accident fatalities. He expressed that there exists an inverse relationship between them in both developed and developing countries. Valli [6] suggested accident prediction models for India and metropolians of India by benefiting from Smeed and Andreassen accident models, and predicted the number of accidents, fatalities and injuries for 2007 and 2010. Zegeer [7] developed an accident model by using traffic, road geometry and terrain condition data, which were assumed to be effective parameters. Chakrobort and Roy [8] conducted a study to determine the road safety level in Kolkata, India based on Smeed model. Junus *et al.* [9] compared two-time series methods that are time series regression and structural times of modelling road accidents in Penang and their results showed that both models have a different view in terms of the relationship and the factor contributing to road accidents.

Recently, artificial intelligence methods have been used in engineering applications including model developments. Differential Evolution Algorithm were applied to PID controllers for time delay systems by Shahri *et al.* [10]. Zellagui *et al.* [11] applied Differential Evolution Algorithm for multi-objective optimization of power transformer design as well. Artificial Neural Network (ANN) which is another of artificial intelligence methods used for estimation of delay and vehicle stops at signalized intersections by Doğan *et al.* [12]. Mussone *et al.* [13] proposed a model utilizing Artificial Neural Network method for the analysis of vehicle accidents in Milan, Italy. Akgungor and Dogan [14] developed accident prediction models for Turkey with ANN and nonlinear regression approaches. The same researchers performed another study using ANN and Genetic Algorithm (GA) to develop accident models for the city of Ankara. The results showed that the performance of ANN models was better than that of GA models [15]. Same authors proposed various ANN accident models based on Smeed and Andreassen approaches for three metropolitan cities of Turkey, and evaluated the performances of the models in terms of errors measurements [16]. Codur and Tortum [17] developed an accident prediction model using road and traffic parameters with ANN approach. The study result showed that the degree of vertical curvature is the most important parameter that affects the number of accidents on highways.

In this study, three accident prediction models for Turkey were developed based on the population and the number of vehicle using DEA. Linear, exponential and semi-quadratic forms were used for the accident models. The estimates of the models were statistically compared to real data. The best fit model was selected to forecast the number of traffic accidents for the future safety studies.

2. Methodology

2.1 Differential Evolution Algorithm

Lately, many researchers have used population-based meta-heuristics search algorithms to solve engineering problems because their performances are usually better than that of traditional methods. DEA proposed by Storn and Price [18] is one of the meta-heuristic algorithms. The basic principle of DEA is quite similar to GA. Four basic operators run in DEA. These are initial population, mutation, crossover and selection. The main difference between DEA and GA is in the mutation operator. The mutation process based on the difference of randomly selected vectors in DEA improves the performance of algorithm and allows to be found the local minima by ignoring the initial parameter values. All operators in DEA are not applied to the whole population one by one as GA. DEA requires less control parameters and can be encoded with less lines. These advantages of algorithm make it more popular and preferred approach to solve the many kinds of problems.

The number of input variables (D) in the initial population determine the population size and the number of chromosomes in the population must be more than 3 [19]. Initial population is obtained by random distribution between 0 and 1. Mathematical form of initial population is shown by Eq.1.

$$X_{j,i}^{(0)} = X_j^{min} + rand_j(0,1) \cdot (X_j^{max} - X_j^{min}) \quad (1)$$

where X_j^{max} and X_j^{min} are the upper and lower bounds of the jth parameter, $i = 1 \dots N_p$ (number of population) and $j = 1 \dots D$ (number of parameters in fitness function).

Mutation operator uses the difference between members of population and requires three randomly selected chromosomes. Thus the mutant vector is obtained. Mathematical form of mutation is shown by Eq.2.

$$X_i'^{(G)} = X_a^{(G)} + F(X_b^{(G)} - X_c^{(G)}) \quad i = 1 \dots N_p \quad (2)$$

in which $X_i'^{(G)}$ is the mutant vector, $X_a^{(G)}$ is the base vector, G is generation number, F is the scaling constant, $X_b^{(G)}$ and $X_c^{(G)}$ are random vectors to produce the difference vector.

There are a lot of strategies in mutation and they are expressed in DGA/x/y/z form. Here, x refers to the vector employed to create mutant vectors, y is the number of difference vectors used in the mutation process and z is the crossover scheme utilized in the crossover operation. The aim of the crossover process is crossover between the target vector and mutant vector. Therefore, trial vector is obtained. Mathematical form of crossover is shown by Eq.3.

$$X_i''^{(G)} = \begin{cases} X_i'^{(G)} & \text{if } rand_j(0,1) \leq C_r \text{ or } j = j_{rand} \\ X_i^{(G)} & \text{otherwise} \end{cases} \quad (3)$$

C_r is the crossover probability between 0 and 1.

Crossover is executed according to selected probability distribution [20].

The last step of DEA is to select the best population between target and trial vectors. Expression of this is shown by Eq. 4.

$$X_i^{(G+1)} = \begin{cases} X_i^{''(G)} & \text{if } f(X_i^{''(G)}) \leq f(X_i^{(G)}), i = 1 \dots Np \\ X_i^{(G)} & \text{otherwise} \end{cases} \quad (4)$$

The values of control parameters. F, Cr and Np, should be selected in optimum ranges as Storn suggested. This is an important step for the model performance [18].

2.2 Development of the models

To develop the accident models, necessary data were obtained from Turkish Statistical Institute (TSI, 2014). Three different mathematical models, namely linear, exponential and semi-quadratic forms, given in Eqs. [5-7] respectively, proposed to estimate the number of accident in Turkey. The purpose of using the models in three different forms is to establish the relationship between the number of accidents and the parameters affecting them. Therefore, linear and exponential forms, which are easy to use and conform to the distribution of independent parameter values, are preferred. In addition, since independent input variables interact with each other, it is preferred in this study to demonstrate the performance of the semi-quadratic form, which is in a more complex form.

$$A = w_1 * x_1 + w_2 * x_2 + w_3 \quad (5)$$

$$A = w_1 * x_1^{w_2} + w_3 * x_2^{w_4} + w_5 \quad (6)$$

$$A = w_1 * x_1 + w_2 * x_2 + w_3 \sqrt{x_1 * x_2} + w_4 \quad (7)$$

in which x_1 is the number of population and x_2 is the number of vehicles. W_i is the corresponding weighting factors.

The coefficients of the models given in Table 1 are obtained by using DEA.

Table 1: The coefficient of models

Linear	Exponential	Semi-Quadratic
	$w_1 = 48,6580$	$w_1 = 0,2659$
$w_1 = 0,03524$	$w_2 = 0,5990$	$w_2 = -0,3209$
$w_2 = 0,04756$	$w_3 = 149521,2$	$w_3 = 0,0275$
$w_3 = -2984892$	$w_4 = 0,2440$	$w_4 = -4966102$
	$w_5 = -12388411$	

For verification, the data were randomly separated into training and test. For this purpose, 11 data were used for training remaining data were utilized for test

3. Result and discussion

3.1 Comparison of the models

The performances of models were evaluated in terms of error criteria which are root mean square errors (RMSE), mean absolute percentage errors (MAPE) and coefficient of determination (R^2) defined in Eqs. 8-10.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (A_{Observed} - A_{estimated})^2} \quad (8)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_{Observed} - A_{estimated}}{A_{Observed}} \right| * 100 \quad (9)$$

$$R^2 = 1 - \left[\frac{\sum_{i=1}^n (A_{Observed} - A_{estimated})^2}{\sum_{i=1}^n (A_{Observed} - A_{mean})^2} \right] \quad (10)$$

As seen from Table 2, the R^2 values of all models for training and test are close to each other. On the other hand, the RMSE and MAPE values of linear model for training and test are better than those of exponential and semi-quadratic models except the RMSE of training in semi-quadratic model. Therefore, developed linear accident model is quite reliable and can be used an efficient tool for future estimations.

Table 2: A compression of errors in developed models for training and test

	Linear		Exponential		Semi-Quadratic	
	Training	Test	Training	Test	Training	Test
RMSE	72124,57	94521,93	82785,42	114327,34	70717,01	115881,58
MAPE	6,37	13,05	7,55	13,56	8,02	15,30
R^2	0,9479	0,9419	0,9530	0,9422	0,9601	0,9330

The estimates of accidents by developed models are graphically compared against the observed values in Fig. 1. As the figure indicates all the developed models catch the increasing trend of traffic accidents.

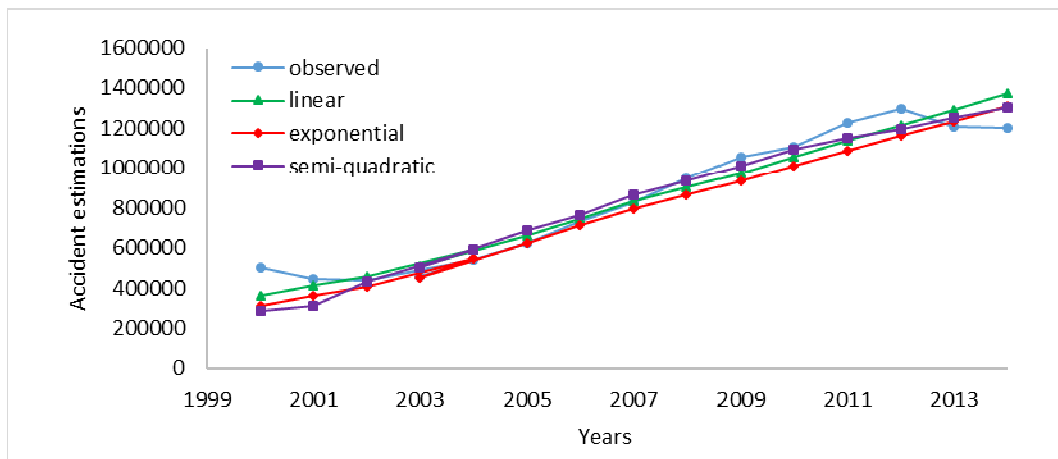


Figure 1: Compression of the estimated and observed accidents in Turkey

3.2 Future estimates of traffic accidents for Turkey

In this section of the study, a scenario is considered to estimate the number of accidents until 2024. According to the projection of TSI, the population of the country will close to 85 million in 2024. The vehicle ownership rate is expected to reach 0.4 by 2024. Thus the number of the motor vehicles will reach to approximately 34 million. This rate is consistent with the average vehicle ownership rate in European countries. The reader is referred to Table 3 for data on the population and motor vehicle data used in this scenario. Linear model estimates are tabulated in Table 4. The study results show that the number of the accident in Turkey will be 2 245 567 in 2024.

Table 3: Population and vehicle predictions

Year	Future Estimates	
	Population	Number of Vehicles
2015	78 151 750	19 537 938
2016	78 965 645	21 320 724
2017	79 766 012	22 334 483
2018	80 551 266	24 165 380
2019	81 321 569	25 209 686
2020	82 076 788	27 085 340
2021	82 816 250	28 985 688
2022	83 540 076	30 909 828
2023	84 247 088	32 013 893
2024	84 936 010	33 974 404

Table 4: Linear model estimates

Year	Accident Estimates
2015	1415052
2016	1516466
2017	1590167
2018	1691915
2019	1765263
2020	1867159
2021	1969176
2022	2071288
2023	2143734
2024	2245467

4. Conclusion

This study presents an application of DEA to estimate the number of traffic accidents in Turkey for the next decade. Traffic accident prediction models have developed by using a fifteen-year historical data covering the years between 2000 and 2014. In the model development, the population and number of motor vehicles are considered as model parameters. Three different types of accident models incorporating a linear, an exponential and a semi-quadratic form of mathematical expressions have developed. A comparative study is performed among developed accident models and observed data. The model results indicate that all of them catch the increasing trend of traffic accidents. Therefore, any of the developed model is used alternatively for accident predictions.

All developed models are statistically compared in terms of error criteria, such as RMSE, MAPE and R^2 , for both training and test stage. The results show that the performance of linear accident model is relatively better than that of others. Hence, linear accident model is selected as the best model for the future estimates. Then, the performance of the linear accident model is evaluated with a scenario. This scenario covers ten-year period from 2015 to 2024. According to the scenario, the population and number of vehicles in Turkey will reach nearly 85 and 34 million in 2024, respectively. During the same period the number of accident will steadily increase and go up from 1 415 052 to 2 245 467 if necessary preventions for traffic safety are not taken.

In this study, for the sake of simplicity and practical usage two common parameters, the population and number of vehicle, were selected. For further studies, other parameters which are effected on traffic accidents could be employed to improve the performance of the models.

References

- [1] World Health Organization. (2011). Global Plan fort the Decade of Action for Road Safety 2011-2020. Accessed 09 December 2015, from http://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf

- [2] Turkish Statistical Institute. (2015). Traffic Accident Statistics 2000-2014. Accessed 28 November 2015, from <http://www.tuik.gov.tr>
- [3] Smeed, R. J. (1949). Some Statistics Aspects of Road Safety Research. *Journal of the Royal Statistical Society. Series A, Part I*, 1-34.
- [4] Andreassen, D. C. (1985). Linking Deaths with Vehicles and Population, *Traffic Engineering & Control*. 26 (11), 547-549.
- [5] Mekky, A. (1985). Effect of Rapid Increase in Motorization Levels on Road Fatality Rates in Some Rich Developing Countries. *Accident Analysis & Prevention*. 17 (2), 101-109.
- [6] Valli, P. P. (2005). Road Accident Models for Large Metropolitan Cities of India. *IATSS Research*. 29 (1), 57-65.
- [7] Zegeer, C. V. & Deacon J. A. (1985). Effect of Lane Width, Shoulder Width, and Shoulder Type on Highway Safety. In: Relationship between Safety and Key Highway Features. *State of Art Report 6*. Transportation Research Board Washington D.C.
- [8] Chakrobort, S. & Roy, S. K. (2005). Traffic Accident Characteristics of Kolkata. *Transport and Communications Bulletin for Asia and the Pacific*. 74, 75-86.
- [9] Junus, N. W. M., İsmail, M. T. & Arsad, Z. (2015). Predicting Penang Road Accidents Influences: Time Series Regression Versus Structural Time Series. *Indian Journal of Science and Technology*. 8 (30).
- [10] Shahri, M.E., Balochian, S., Balochian, H. & Zhang, Y. (2014). Design of Fractional – order PID Controllers for Time Delay Systems Using Differential Evolution Algorithm. *Indian Journal of Science and Technology*. 7 (9), 1307-1315.
- [11] Zellagui, M., Hassan, H. A. & Abdelaziz, A. Y. (2017). Non-dominated sorting gravitational search algorithm for multi-objective optimization of power transformer design. *Engineering Review*. 37 (1), 27-37.
- [12] Doğan, E., Akgüngör, A. P. & Arslan, T. (2016). Estimation of delay and vehicle stops at signalized intersections using artificial neural network. *Engineering Review*. 36 (2), 157-165.
- [13] Mussone, L., Ferrari, A. & Oneta, M. (1999). An Analysis of Urban Collision Using an Artificial Intelligence Model. *Accident Analysis and Prevention*. 31 (8), 705-718.
- [14] Akgüngör, A. P. & Doğan, E. (2008). Estimating Road Accidents of Turkey Based on Regression Analysis and Artificial Neural Network Approach. *Advances in Transportation studies an International Journal Section A* 16. 11-22.
- [15] Akgüngör, A. P. & Doğan, E. (2009). An Artificial Intelligent Approach to Traffic Accident Estimation: Model Development and Application. *Transport*. 24 (2), 135-142.
- [16] Akgüngör, A. P. & Doğan, E. (2009). An Application of Modified Smeed, Adapted Andreassen and Artificial Neural Network Accident Models to Three Metropolitan Cities of Turkey. *Scientific Research and Essays*. 4 (9), 906-913.
- [17] Codur, M. Y. & Tortum, A. (2015). An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey. *Promet-Traffic & Transportation*. 27 (3), 217-225.
- [18] Storn, R. & Price, K. (1997). Differential Evolution – A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. *Journal of Global Optimization*. 11 (4), 341-359.

- [19] Storn, R. (2008). Differential Evolution Research – Trends and Open Questions Advances in Differential Evolution, *Springer-Verlag Berlin Heidelberg*. Chakraborty, U.K.
- [20] Karaboğa, D. (2004). Artificial Intelligence Optimization Algorithms (in Turkish). *Atlas Yayın Dağıtım*. İstanbul.

