FUZZY RENEWAL THEORY ABOUT FORECASTING MISTAKES DONE BY A LOCOMOTIVE DRIVER: A SERBIAN RAILWAY CASE STUDY

Milan Marković1, Norbert Pavlović2, Miloš Ivić3
Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, Belgrade, Serbia
E-mails: 1milan@sf.bg.ac.rs (corresponding author); 2norbert@sf.bg.ac.rs; 3m.ivic@sf.bg.ac.rs

Submitted 31 January 2011; accepted 8 March 2011

Abstract. The human factor is one of the most dominant causes of railway accidents. For example, human impact appears to be the main reason for 44% of railway accidents in the Republic of Serbia. Thus, a remarkable effort is undertaken to investigate human factors. Therefore, plenty of researchers have analyzed a human influence on railway accidents. This paper develops a model for forecasting the number of railway accidents caused by the human factor. The proposed model is based on the renewal theory and assumes that working time between the faults of a locomotive driver has exponential distribution (or another Erlang distribution of a higher order) characterized by parameter $\lambda$ that is treated as a fuzzy dependant variable and considered as a function of job complexity, the exposure of locomotive drivers (i.e. time spent in driving) and a tendency of locomotive drivers to make mistakes. The application of the model to the population of 777 Serbian railway locomotive drivers provided encouraging results in predicting the number of railway accidents.

Keywords: traffic, railway accident, human factor, fuzzy theory, forecast, locomotive driver.

1. Introduction
The aim of the present paper is to develop the model for forecasting the number of traffic accidents caused by locomotive drivers considering job complexity, the exposure of locomotive drivers and the tendency of locomotive drivers to make mistakes. This work is based on the renewal theory, develops a model for predicting the number of railway accidents and assumes that working time between two failures of a single locomotive driver has exponential distribution (or another Erlang distribution of higher order) characterized by parameter $\lambda$. The model suggested in this article treats parameter $\lambda$ as a dependent fuzzy value. We consider parameter $\lambda$ to be a function of job complexity, the exposure of locomotive drivers and the tendency of locomotive drivers to make mistakes.

The paper includes five sections. The following section shows the results of the previously conducted research. The third section defines the model in which the parameter of the reliability function is considered as a fuzzy value. The fourth section applies the model for the case study based on real statistical data related to Serbian Railways. Finally, conclusions are drawn.

2. Previous Research
Research on the human factor in railway accidents can be divided into two groups. The first group covers the papers that tend to identify precautionary measures that could help with reducing the number of accidents (Harms-Ringdahl 2009; Reer 2008; Baysari et al. 2009; McLeod et al. 2005; Vanderhaegen 2001; Santos-Reyes, Beard 2009).

This work belongs to the second group of papers focused on the development of models for forecasting the number of railway accidents caused by the human factor. We focus on statistical data analysis. Related research such as (Evans 2003; Marković et al. 2007) enables the persons involved in traffic safety to be warned about potential problems before they may arise. However, the aforementioned research does not identify the type of precautionary measures to be taken.

The authors of the paper (Markovic et al. 2007) observe a locomotive driver as an element of a technical system and prove that the renewal process is Poisson’s. They emphasize that the number of mistakes done by a locomotive driver until time $t$ can be described by ran-
dom variable $N(t)$ having Poisson probability distribution with parameter $\lambda$:
\[
P_n(t) = P\left[ N(t) = n \right] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}
\]
for $n = 1, 2, 3, ...$
\[
(1)
\]
where: $N(t)$ – represents the number of mistakes that a locomotive driver made until time $t$, $\lambda$ – denotes the parameter of Poisson distribution; $t$ – is observed time.

Equation (1) is general and enables forecasting the total number of mistakes. However, the problem arises in estimating parameter $\lambda$. It can be determined based on statistical analysis, but the problem is encountered when no adequate historical data are available or when some changes in the railway system are made. This paper develops a method of determining parameter $\lambda$ for the two aforementioned cases.

3. The Model

We statistically analyzed failure occurrence $\lambda$ on Serbian railways. The aim of the conducted investigation was to examine the possibility of applying equation (1). The received results showed a significant dispersion of parameter $\lambda$, implying that there were significant inter- and intra-network differences in reliability among locomotive drivers. These differences can be explained by varying working conditions, safety systems, railway track parameters and states, railway traffic intensities, traction systems, etc. All these factors influence the complexity of jobs performed by locomotive drivers. Differences in work reliability among locomotive drivers working in the same railway network can be explained by different psychophysical characteristics and living conditions as well as by different driving exposures. For these reasons, we consider parameter $\lambda$ as a function of various independent parameters including the complexity of an assignment, driving exposure and the tendency of locomotive drivers to make mistakes. Since no adequate statistical database related to Serbian railways exists, we apply the fuzzy set theory and develop a fuzzy system.

The fuzzy set theory is a convenient mathematical instrument for researching phenomena characterized by uncertain consequences, subjective valuation and ambiguity (Teodorović, Vukadinović 1998; Jović, Popović 2001), particularly when no adequate database is available. A number of published papers deal with the fuzzy set theory. These papers have shown a possibility of applying the above introduced theory to solve practical traffic problems (Vukadinović, Popović 2008; Teodorović, Vukadinović 1998; Teodorović 1999).

The following entry fuzzy variables were introduced:

- $X_1$ – the tendency of a locomotive driver to make a mistake (i.e. to cause a traffic accident), defined as an unnamed value in the interval $[0, 1]$; the number is assigned based on the number of mistakes taking into consideration the complexity of work tasks and the time spent while driving;
- $X_2$ – the number of hours per year spent while driving;
- $X_3$ – the complexity of tasks performed by a locomotive driver.

Fuzzy variables $X_1$, $X_2$ and $X_3$ represent entry variables. Exit variable $\lambda$ is determined by the values of entry variables. Thus, parameter $\lambda$ is a fuzzy variable that describes the intensity of the occurrence of mistakes for one locomotive driver.

The fuzzy variables of the entry were considered as fuzzy sets (small, medium and large):

We defined fuzzy membership functions for slight, medium and strong tendency as linear functions:

\[
\mu_{\text{small}}(X_1) = \begin{cases} 
1 - 2.5x_1 & \text{for } 0 \leq x_1 \leq 0.4; \\
0 & \text{for } x_1 \geq 0.4;
\end{cases}
\]

(2)

\[
\mu_{\text{medium}}(X_1) = \begin{cases} 
10x_1 & \text{for } 0 \leq x_1 \leq 0.1; \\
1.2 - 2x_1 & \text{for } 0.1 \leq x_1 \leq 0.6; \\
0 & \text{for } x_1 \geq 0.6;
\end{cases}
\]

(3)

\[
\mu_{\text{large}}(X_1) = \begin{cases} 
0 & \text{for } x_1 \leq 0.29; \\
\frac{200}{81}x_1 - \frac{59}{81} & \text{for } 0.29 \leq x_1 \leq 0.7; \\
1 & \text{for } x_1 \geq 0.7.
\end{cases}
\]

(4)

A graphical scheme for fuzzy membership functions is shown in Fig. 1.

We defined fuzzy membership functions for limited, medium and great driving exposures as linear functions:

\[
\mu_{\text{small}}(X_2) = \begin{cases} 
1 & \text{for } x_2 \leq 100; \\
\frac{10}{9} - \frac{x_2}{900} & \text{for } 100 \leq x_2 \leq 1000; \\
0 & \text{for } x_2 \geq 1000;
\end{cases}
\]

(5)

\[
\mu_{\text{medium}}(X_2) = \begin{cases} 
0 & \text{for } 0 \leq x_2 \leq 300; \\
\frac{x_2}{5} - \frac{3}{500} & \text{for } 300 \leq x_2 \leq 800; \\
\frac{7}{3} - \frac{x_2}{600} & \text{for } 800 \leq x_2 \leq 1400; \\
0 & \text{for } x_2 \geq 1400;
\end{cases}
\]

(6)

\[
\mu_{\text{large}}(X_2) = \begin{cases} 
0 & \text{for } x_2 \leq 750; \\
\frac{x_2}{550} - \frac{15}{11} & \text{for } 750 \leq x_2 \leq 1300; \\
1 & \text{for } x_2 \geq 1300.
\end{cases}
\]

(7)

A graphical scheme for fuzzy membership functions is shown in Fig. 2.

We define fuzzy variable $X_3$ in the interval $[1, 5]$ of the whole number set:

1. a railway track with modern systems of the leading train;
2. a two line track with modern signal units, an installed track and locomotive beacons;
3. an one-track line with modern signal units, an installed track and locomotive beacons;
4. a two track line with a safety system and no dependence between signal showing and switch location; 
5. an one-track line with a safety system no dependence between signal showing and switch location.

We define fuzzy membership functions for decreased, medium and increased job complexities as:

\[
\mu_{\text{small}}(x_3) = \begin{cases} 
1 & \text{for } x_3 = 1; \\
0.75 & \text{for } x_3 = 2; \\
0.50 & \text{for } x_3 = 3; \\
0.25 & \text{for } x_3 = 4; \\
0 & \text{for } x_3 = 5; 
\end{cases} 
\]  
(8)

\[
\mu_{\text{medium}}(x_3) = \begin{cases} 
0 & \text{for } x_3 = 1; \\
0.50 & \text{for } x_3 = 2; \\
0.50 & \text{for } x_3 = 3; \\
0 & \text{for } x_3 = 4; \\
0 & \text{for } x_3 = 5; 
\end{cases} 
\]  
(9)

\[
\mu_{\text{large}}(x_3) = \begin{cases} 
0.50 & \text{for } x_3 = 3; \\
0.25 & \text{for } x_3 = 4; \\
1 & \text{for } x_3 = 5. 
\end{cases} 
\]  
(10)

A graphical scheme for fuzzy membership functions is shown in Figs 3–5.

Exit parameter \( \lambda \) is also a fuzzy variable. We defined fuzzy membership functions for irrelevant, medium, relevant and highly relevant parameter \( \lambda \) as linear functions:

\[
\mu_{\text{small}}(\lambda) = \begin{cases} 
1 & \text{for } 0 \leq \lambda \leq 0.1; \\
\frac{3}{2} & \text{for } 0.1 \leq \lambda \leq 0.3; \\
0 & \text{for } \lambda \geq 0.3; 
\end{cases} 
\]  
(11)

\[
\mu_{\text{medium}}(\lambda) = \begin{cases} 
7.69\lambda & \text{for } 0 \leq \lambda \leq 0.13; \\
\frac{40}{27} - \frac{100}{27}\lambda & \text{for } 0.13 \leq \lambda \leq 0.4; \\
0 & \text{for } \lambda \geq 0.4; 
\end{cases} 
\]  
(12)

\[
\mu_{\text{large}}(\lambda) = \begin{cases} 
0 & \text{for } 0 \leq \lambda \leq 0.2; \\
\frac{10}{3} & \text{for } 0.2 \leq \lambda \leq 0.5; \\
\frac{7}{5} - 5\lambda & \text{for } 0.5 \leq \lambda \leq 0.7; \\
0 & \text{for } \lambda \geq 0.7; 
\end{cases} 
\]  
(13)

Fig. 1. Fuzzy membership functions for slight, medium and strong tendencies

Fig. 2. Fuzzy membership functions for limited, medium and great exposures

Fig. 3. Fuzzy membership functions for decreased job complexity

Fig. 4. Fuzzy membership functions for medium job complexity

Fig. 5. Fuzzy membership functions for increased job complexity
The established functions are shown in Fig. 6.

The base of rules is a part of the fuzzy system and consists of 27 rules shown in Table 1. The quantification of these rules is carried out by an approximation algorithm.

For example, rule 5 explains that if \( X_1 \) is low, \( X_2 \) is medium and \( X_3 \) is medium, then \( \lambda \) is medium. The expression 'if \( X_1 \) is low, \( X_2 \) is medium and \( X_3 \) is medium' shows the fuzzy phrase defined in group \( X_1 \times X_2 \times X_3 \). If we denote this fuzzy phrase by \( N \), then, the fuzzy membership function of this phrase is:

\[
\mu_{\text{very large}}(\lambda) = \begin{cases} 
0 & \text{for } 0 \leq \lambda \leq 0.5; \\
2\lambda - 1 & \text{for } 0.5 \leq \lambda \leq 1.
\end{cases}
\] (14)

Therefore, rule 5 represents fuzzy implication that can be written as 'if \( N \), then \( \lambda \) is medium'. This fuzzy implication represents a fuzzy phrase defined for group \( X_1 \times X_2 \times X_3 \times \lambda \). We define this fuzzy phrase as \( M \). The fuzzy membership function for this phrase is:

\[
m_M(\lambda) = \min\left\{m_N(X_1, X_2, X_3), m_\text{medium}(\lambda)\right\}.
\]

We define fuzzy phrases for all 27 algorithm rules taken from Table 1 as \( M_n \) (\( n = 1, 2, ..., 27 \)). These fuzzy phrases are joined by 'OR' and represents a fuzzy sentence. We denote the fuzzy sentence by \( R \). The fuzzy membership function for fuzzy sentence \( R \) is:

\[
\mu_R = \max\left\{\mu_{M_1}(X_1, X_2, X_3, \lambda), \mu_{M_2}(X_1, X_2, X_3, \lambda), ..., \mu_{M_{27}}(X_1, X_2, X_3, \lambda)\right\}.
\]

Next, we assign a degree of membership to each possible value of exit variable \( \lambda \) depending on entry variables \( X_1, X_2 \) and \( X_3 \) by applying approximation algorithms.

We use one of two rules to choose the value for the exit variable:

- the value that matches the highest degree of the membership;
- adequate coordinate of the gravitation center (centroid).

### Table 1. Base of rules for determining the values of parameter \( \lambda \)

| Rule | Condition | Action |
|------|-----------|--------|
| 1    | If low    | low    |
| 2    | If low    | medium |
| 3    | If low    | high   |
| 4    | If medium | low    |
| 5    | If medium | medium |
| 6    | If medium | high   |
| 7    | If medium | medium |
| 8    | If low    | medium |
| 9    | If low    | high   |
| 10   | If medium | low    |
| 11   | If medium | medium |
| 12   | If medium | high   |
| 13   | If medium | medium |
| 14   | If medium | medium |
| 15   | If medium | high   |
| 16   | If medium | medium |
| 17   | If medium | high   |
| 18   | If medium | high   |
| 19   | If high   | low    |
| 20   | If high   | medium |
| 21   | If high   | high   |
| 22   | If high   | medium |
| 23   | If high   | medium |
| 24   | If high   | medium |
| 25   | If high   | high   |
| 26   | If high   | medium |
| 27   | If high   | high   |

Accident/year \( c_1 \), \( c_2 \), \( c_3 \), \( c_4 \), \( c_5 \)
Defuzzification was performed using the method of the gravitation center within MATLAB. 3D curves shown in Figs 7–9 represent the dependence of two entry values on the exit value.

The surface shown in Fig. 7 represents the relationship between the entry values of tendencies and exposures to exit value $\lambda$. Fig. 7 shows that exposure has little influence on the intensity of accidents for locomotive drivers having slight tendency.

The surface displayed in Fig. 8 represents the relationship between the entry values of tendencies and job complexity to exit value $\lambda$. Fig. 8 indicates that locomotive drivers with a slight tendency to make mistakes have a small number of accidents regardless of the complexity of their work tasks. The influence of work task complexity becomes more significant for locomotive drivers with a stronger tendency.

Finally, the surface in Fig. 9 shows how exposure and job complexity may vary considering $\lambda$. Fig. 9 also shows a more considerable influence of job complexity on the intensity of accidents, except those within the area of great exposure.

4. Results

A two track line equipped with modern signal units, an installed track and a locomotive beacon (job complexity has values 2 and 3) that exist on the railway section 'Lapovo' are taken as a sample. These conditions are given in Table 2.

The results obtained by applying the fuzzy model and shown in Table 3 are related to the reference of the locomotive driver to the railway section 'Lapovo'. Parameter $\lambda$ given by the model and denoted as $\lambda_{sr}$ and $\lambda_{\text{max}}$ was used for forecasting the number of accidents.

Parameter $\lambda_{sr}$ is the median value of parameter $\lambda$ obtained by applying the fuzzy model for different entry values existing in the section 'Lapovo'. The given values $\lambda_{sr}$ are shown in Table 3 taking into account a different diapason of the entry values of tendency, exposure and complexity. Parameter $\lambda_{\text{max}}$ represents the median value of parameter $\lambda$, but for the maximal value of tendency.

The given values $\lambda_{sr}$ are shown in Table 3 taking into account a different diapason of the entry values of tendency, exposure and complexity. Parameter $\lambda_{\text{max}}$ represents the median value of parameter $\lambda$, but for the maximal value of tendency.

| Number of mistakes made by one locomotive driver | Tendency | Exposure [hour/year] | Complexity |
|-----------------------------------------------|----------|----------------------|------------|
| 1 or 2                                        |          | 0 ≤ $S$ ≤ 0.3       | 700 ≤ $I$ ≤ 1100 | 2 and 3   |
| 3 or 4                                        |          | 0.3 < $S$ ≤ 0.7     | 700 ≤ $I$ ≤ 1100 | 2 and 3   |
| 5 or 6                                        |          | 0.7 < $S$ ≤ 1       | 900 ≤ $I$ ≤ 1200 | 3 and 5   |

Work conditions shown in Table 2 are current for 'Lapovo' section. The upper 95% confidence limits on the maximum number of accidents obtained by statistical data are shown in column 8 (Table 3). These values are given by the model described in the paper by Marković et al. (2007). The number of railway accidents given by the developed fuzzy model and values $\lambda_{sr}$ and $\lambda_{\text{max}}$ is shown in columns 9 and 10. The results provided in Table 3 allows making a conclusion that there is an agreement between value $\lambda_{sr}$ and parameter $\lambda$ based on statistical data, which means that there is accordance in the expected number of railway accidents for the observed period. In addition, the correspondence between the number of accidents that will not be exceeded with the probability of 95% given by parameter $\lambda$ from the...
sample and the forecasted number of railway accidents given by fuzzy value \( \lambda_{\text{max}} \) can be noticed. Few differences in Table 3 are due to a small number of locomotive drivers who made more than 4 accidents on ‘Lapovo’ section during the observed time.

The values presented in Table 3 show that parameter \( \lambda \) attained by the fuzzy model corresponds to parameter \( \lambda \) attained from a statistical sample; thus, it can be applied to forecast the number of accidents according to equation (1).

If we vary independent variables \( X_1, X_2, \) and \( X_3 \), we can perceive the influence of the exposure, tendency and complexity of work tasks on the number of accidents that will occur in the future.

### 5. Conclusions

This paper developed a model for forecasting the number of mistakes made by locomotive drivers. The work is based on the renewal theory. We have assumed that work time between two faults for a single locomotive driver has exponential distribution to parameter \( \lambda \). Moreover, we have considered parameter \( \lambda \) to be a function of job complexity, exposure and the tendency of locomotive drivers to make mistakes. We have calibrated and tested the proposed model on the population of 777 locomotive drivers working for Serbian Railways.

We defined fuzzy membership functions for the fuzzy values of tendency and exposure as linear functions. The fuzzy membership function for the fuzzy value of complexity was defined as the whole number set. The base of rules was a part of the fuzzy system and consisted of 27 rules discussed in the paper. The process of defuzzification was performed using the method of the gravitation center within MATLAB.

The developed fuzzy model was applied for a part of Serbian Railway network involving locomotive drivers from ‘Lapovo’ section. Parameter \( \lambda \), presented by the model and denoted as \( \lambda_{\text{sr}} \), and \( \lambda_{\text{max}} \) was used for forecasting the number of accidents.

Parameter \( \lambda_{\text{sr}} \) is the median value of parameter \( \lambda \) obtained by applying the fuzzy model for different entry values existing on section ‘Lapovo’. Parameter \( \lambda_{\text{max}} \) represents the median value of parameter \( \lambda \), but for the maximal value of tendency.

The received results suggest that there was the correspondence between value \( \lambda_{\text{sr}} \) and parameter \( \lambda \) given by statistical data. In addition, we concluded that there was the correspondence between the number of accidents that would not be exceeded with the probability of 95% given by parameter \( \lambda \) from the sample and the forecasted number of railway accidents given by fuzzy value \( \lambda_{\text{max}} \). We observed certain differences that could be explained by a small statistical sample, particularly emphasizing locomotive drivers having more than 4 failures (accidents).

Finally, the case study on Serbian Railways showed that the model provided good results compared to the observed statistical data. The model can be successfully applied to other railway systems to provide adequate calibration. The model seems to be particularly suitable for applications to the railway systems lacking an extensive database.

### Acknowledgment

This paper is a part of the study ‘Research on the Technical-Technological Staff and organizational capacity of Serbian Railways from the Viewpoint of Current and Future European Union Requirements’ and has been financed by The Ministry of Science and Technological Development of the Republic of Serbia.

### References

Baysari, M. T.; Caponecchia, C.; McIntosh, A.; Wilson, J. R. 2009. Classification of errors contributing to rail incidents and accidents: a comparison of two human error identification techniques, Safety Science 47(7): 948–957. doi:10.1016/j.ssci.2008.09.012

Evans, A. W. 2003. Estimating transport fatality risk from past accident data, Accident Analysis and Prevention 35(4): 459–472. doi:10.1016/S0001-4575(02)00024-6

Harms-Ringdahl, L. 2009. Analysis of safety functions and barriers in accidents, Safety Science 47(3): 353–363. doi:10.1016/j.ssci.2008.06.004

Jović, J. J.; Popović, M. M. 2001. Modal split modelling using multicriteria analysis and discrete fuzzy sets, Yugoslav Journal of Operations Research 11(2): 221–233.

Marković, M.; Ivić, M.; Pavlović, N.; Janković, S. 2007. Analysis of simulation model application to forecast the railway workers failures, Yugoslav Journal of Operations Research 17(1): 135–144. doi:10.2298/YJOR0701135M

### Table 3. The forecasted number of railway accidents on Serbian Railways

| Number of mistakes made by one locomotive driver | Number of locomotive drivers in a group | Parameter λ, obtained from the statistical sample | Parameter λ, obtained by the fuzzy model | Observation time [t] | Real number of accidents | Number of accidents that will not be exceeded with the probability of 95% | Number of accidents obtained by the fuzzy model |
|-----------------------------------------------|------------------------------------------|-----------------------------------------------|------------------------------------------|----------------------|------------------------|---------------------------------------------------------------------------------|-----------------------------------------------|
| 1 or 2                                        | 69                                       | 0.174                                         | 0.174                                    | 4                    | 96                     | 112                                                                            | 96                             |
| 3 or 4                                        | 23                                       | 0.424                                         | 0.412                                    | 5                    | 78                     | 92                                                                             | 81                             |
| 5 or 6 or 7                                   | 18                                       | 0.701                                         | 0.721                                    | 6                    | 101                    | 117                                                                            | 104                            |
McLeod, R. W.; Walker, G. H.; Moray, N. 2005. Analysing and modelling train driver performance, *Applied Ergonomics* 36(6): 671–680. doi:10.1016/j.apergo.2005.05.006

Reer, B. 2008. Review of advances in human reliability analysis of errors of commission, Part 1: EOC identification, *Reliability Engineering and System Safety* 93(8): 1091–1104. doi:10.1016/j.ress.2007.07.005

Santos-Reyes, J.; Beard, A. N. 2009. Asystemic analysis of the Edge Hill railway accident, *Accident Analysis and Prevention* 41(6): 1133–1144. doi:10.1016/j.aap.2008.05.004

Teodorović, D. 1999. Fuzzy logic systems for transportation engineering: the state of the art, *Transportation Research Part A: Policy and Practice* 33(5): 337–364. doi:10.1016/S0965-8564(98)00024-X

Teodorović, D.; Vukadinović, K. 1998. *Traffic Control and Transport Planning: a Fuzzy Sets and Neural Networks Approach*. 1st edition. Springer. 387 p.

Vanderhaegen, F. 2001. A non-probabilistic prospective and retrospective human reliability analysis method – application to railway system, *Reliability Engineering and System Safety* 71(1): 1–13. doi:10.1016/S0951-8320(00)00066-0

Vukadinović, S.; Popović, J. 2008. *Matematička Statistika*. Belgrade, Serbia. University of Belgrade (in Serbian).