INCORPORATING DATA-BASED APPROACHES TO DETECT AND SOLVE RECURRING MANAGEMENT PROBLEMS AND CRISES IN OPERATIONAL RELIABILITY AND SAFETY OF ROAD TRANSPORT

Nikolay Georgiev*)

ABSTRACT
Technical advances in vehicle design and production, infrastructure investment and traffic management improvements have lowered transport costs and increased average transport speeds. But many problems remain, for example: city congestion (with all its consequences), traffic accidents, etc. Of course, all involved in transport industry (individuals, transport companies, infrastructure managers, state and municipal authorities, etc.) have their responses to these challenges - people consider extra time for their journeys, companies adapt their routing and timing of operations, infrastructure managers provide regular information regarding traffic flow characteristics.

These factors increase the focus on operational reliability and safety in road transport as a whole, and especially on the need for qualitative decision-making regarding the improvement of their levels. Such decision-making is possible only by a well organized and permanently supported activity of collection, analysis and processing of data in respect of unwanted operational events (infrastructure and car failures, accidents, bad weather conditions, etc.). The possibilities for the application of some well-known statistical approaches and methods for analysis of indices characterizing operational reliability and safety in road transport are discused in the present article.

Key words:
Road Transport, Operational Reliability, Traffic Safety

ABSTRAKT
Technický pokrok v konštrukcii vozidla a výroby, investícií do infraštruktúry a zlepšenie riadenia dopravy znížili náklady na dopravu a zvýšenie priemernej dopravnej rýchlosti. Ale naďalej čelí mnohým problémom, napr: preťaženie mesta (so všetkými

*) Nikolay Dimitrov Georgiev, Assoc. Prof., Ph.D, Higher School of Transport – Sofia 1574, „Geo Milev“ Str, No 158, email: safetyniky@mail.com

jeho následkami), dopravné nehody, atď. Samozrejme, že všetky zúčastnené strany v odvetví dopravy (jednotlivci, dopravné podniky, manažéri infraštruktúry, štátne a obecné úrady, atď) majú schopnosť reagovať na tieto problémy - ľudia považujú za ďalší čas pre svoje cesty, spoločnosť prispôsobuje svoje trasy a načasovania operácií, manažéri infraštruktúry poskytujú pravidelné informácie o vlastnostiach dopravného prúdu.

Tieto faktory zvyšujú dôraz na prevádzkovú spoľahlivosť a bezpečnosť v cestnej doprave ako celku, a najmä na potrebu kvalitatívne rozhodovanie, pokiaľ ide o zlepšenie ich úrovne. Takéto rozhodovanie je možné iba v dobre organizovanej a trvalej podporovanej činnosť zberu, analýzy a spracovania údajov vo vzťahu ku nežiaducich prevádzkových udalostí (infraštruktúry a poruchy automobilu, nehody, zlé poveternostné podmienky, atď). Možnosti pre uplatnenie niektorých známych štatistických prístupoch a metódach pre analýzu ukazovateľov charakterizujúcich prevádzkovú spoľahlivosť a bezpečnosť v cestnej doprave, sú diskutované v tomto článku.

Kľúčové slová:
Cestná doprava, prevádzková spoľahlivosť, bezpečnosť dopravy

1. INTRODUCTION

Undoubtedly, transport is at a turning point. On one hand the role of transport for the progress of modern society is more than ever vital. But on the other, it is connected with many daily problems, crisis situations and additional requirements.

Technical advances in vehicle design and production, infrastructure investment and traffic management improvements have lowered transport costs and increased average transport speeds. But many problems remain, for example: city congestion (with all its consequences), traffic accidents, etc. Of course, all involved in transport industry (individuals, transport companies, infrastructure managers, state and municipal authorities, etc.) have their responses to these challenges - people consider extra time for their journeys, companies adapt their routing and timing of operations, infrastructure managers provide regular information regarding traffic flow characteristics.

These factors increase the focus on operational reliability and safety in road transport as a whole, and especially on the need for qualitative decision-making regarding the improvement of their levels. Such decision-making is possible only by a well organized and permanently supported activity of collection, analysis and processing of data in respect of unwanted operational events (infrastructure and car failures, accidents, bad weather conditions, etc.).

This article considers possibilities for the application of some well-known statistical approaches and methods for analysis of indices characterizing operational reliability and safety in road transport.

2. Key fields to improve operational reliability (and safety) that require analysis of statistics

Like many countries, Bulgaria has to solve lots of problems in the area of operational reliability and safety in road transport. They all require an attentive assessment that could be successively achieved by analysis of statistical data. The most important of these problems are:

- Identification of the causal factors involved in adverse transport events (e.g. accidents), including their location of occurrence and features;

- Identification and assessment of the influence of traffic flow and road infrastructure characteristics on reliability of the transport service;

- Providing an early consideration of operational reliability and safety in all investment programs in respect of road infrastructure construction and reconstruction projects;

- Identification of safety problems of some special road users (such as pedestrians, bicyclists, carriers of hazardous materials, etc.) within the technical exploitation of road transport;

- Optimization of urban transportation planning.

3. Selection and statistical analysis of appropriate indicators for assessment of operational reliability and safety

3.1. Travel time reliability

Travel time is a widely used term in the field of transport and measures the time that it takes for a given transport vehicle to pass over the distance of a transport route. Travel time is a variable and its current value depends on the momentary negative influence of many operating factors. This is why the estimation of travel time is of increasing importance to all users and managers of a given transport system. In practice, the reliability of estimates of travel time regarding a certain transport route is more important for passengers, shippers and transport managers than the travel time itself.

Recently, reliability is extensively used in many engineering fields, which by definition means the probability of successful accomplishment one job under certain conditions. Introduced into the field of transport exploitation, the travel time reliability stands for the probability of a vehicle arriving at its destination within a given time period in the current operating environment. Because the operating environment changes day to day and/or over different periods of the day, the travel time reliability should be assessed in a historical sense (processing of historical data): examination of travel time distribution and specific statistics (its mean, median, standard deviation, variance, etc.). There are five key measures used to assess the variability of travel time in the course of different months, days and day periods [4]. They are as follows:

-90 th (or 95 th) percentile travel time – it represents the worst travel times for given transport route. In other words, it indicates that 90 (or 95) percent of travel times for that route are shorter than the 90 th (or 95 th) percentile travel time - i.e. the probability that travel time is shorter than that percentile is 0,90 (0,95).

-Buffer index (β) – represents the extra time that a hypothetical traveller should add to his/her average travel time to have on-time arrival. The buffer index is calculated by the following expression:

$$\beta = \frac{T_{95} - m_T}{m_T}, \quad [\%] \tag{1}$$

where:

T_{95} - 95 th percentile travel time;

m_T - average (mean) travel time.

As a travel time reliability measure, Buffer index allows computing the indicator **Buffer time (**T_β**):**

$$T_\beta = \beta m_T . \quad [\text{min}] \tag{2}$$

The traveller should have total time for the trip $T_{trip} = m_T + T_\beta$ in order to ensure on-time arrival 95 percent of the time.

-Travel time index ($\beta_{travel\ time}$) – represents the comparison between travel time during peak hours and that time under free-flow conditions:

$$\beta_{travel\ time} = \frac{m_T}{T_{free\ flow}}, \tag{3}$$

where:

$T_{free\ flow}$ – free-flow travel time.

-Planning time index ($\beta_{planning\ time}$) – represents the total planned time needed for an on-time arrival 95% of the time:

$$\beta_{planning\ time} = \frac{T_{95}}{T_{free\ flow}}. \tag{4}$$

Planning time index allows computing the indicator **Planning time (**$T_{planning}$**)**:

$$T_{planning} = \beta_{planning\ time} T_{free\ flow} . \quad [\text{min}] \tag{5}$$

-Percentage of travels "on-time" – percentage of the total travels, fulfilled within a previously designed travel time (e.g. average travel time increased with 10 or 20 percent).

The estimation of mentioned above measures is the main difficulty connected with the assessment of travel time reliability. Their calculation could be based on continuous probe vehicle data, point-based detector data, intentional studies, or simulation.

All measures of travel time reliability are very important indicators which could be used for assessment of the momentary state of a transport system [1], [5]. For instance, their predictability can be a powerful tool to assess and improve the quality of bus services (planning and operation).

Operational reliability of bus routes is one of the most important components of bus service quality within an urban area. It can be defined as the accordance of implementation of a given bus route with a pre-determined transportation plan (timetable) - designed in compliance with specific criteria and operational indicators. Thus, *punctuality* is probably the most widely used term when it comes to the reliability of bus service and is associated with the timely arrival/departure of buses at/from stops within the respective route. Because of the actual influence of some affecting factors (such as weather conditions, accidents, traffic conditions, etc.) there always is some variability associated with the travel time of a bus route (i.e.

punctuality). The greater this variability, the less consistent is the bus route with the pre-determined requirements. In this sense, determining the probability of failure of a timetable (probability of delays in some street sections) is of great importance.

The collection and processing of real data for the travel time (as a random variable) of a bus route (and on this base determination of its distribution function) is a most realistic way to assess punctuality of the bus service. If there is a lack of such data it is possible to use simulation.

As a very flexible class of probability density functions, Gamma distribution is convenient for modelling (simulating) a wide range of random processes including travel time. It has two parameters: *shape parameter* (r) and *scale parameter* (λ). The parameter r represents the deviation of real travel time through a bus route from the designed travel time (in accordance with the timetable). Thus, by using certain data regarding a real bus service and with the help of specialized simulation software it is possible to model bus travel time as Gamma distribution. The simulation could be a base to assess the reliability of designed bus timetable (probability of delay along different sections).

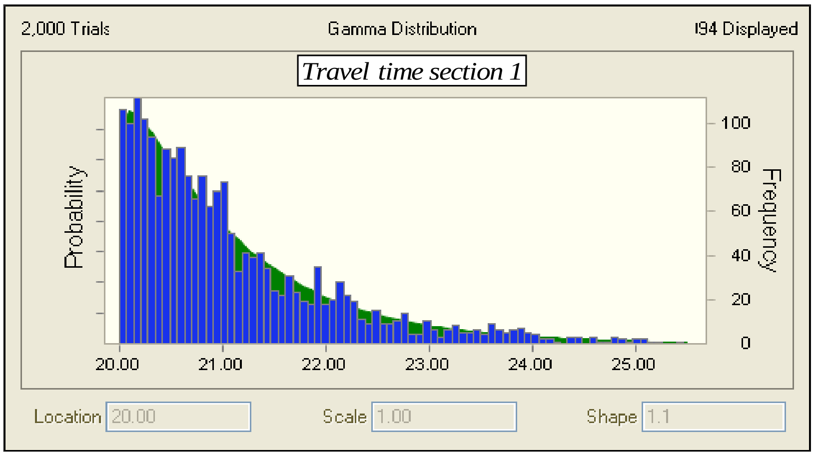

Fig.1.Simulation distribution of bus travel time through secton 1

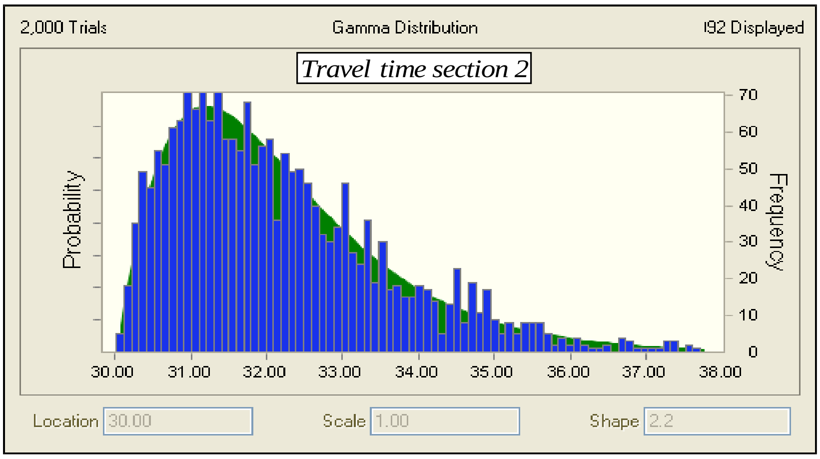

Fig.2.Simulation distribution of bus travel time through secton 2

Let us suppose that a bus route consists of two sections and its length (and structure) is composed of the lengths (and stops) of the individual sections. The overall travel time of buses through the route (T) will be a combination of travel times along the two sections which according to previously designed bus timetable let be: $T_{1,timetable} = 20\ min$ and $T_{2,timetable} = 30\ min$. The determination of probability of delay of buses when travelling through first and second section (also through the route as a whole) is of

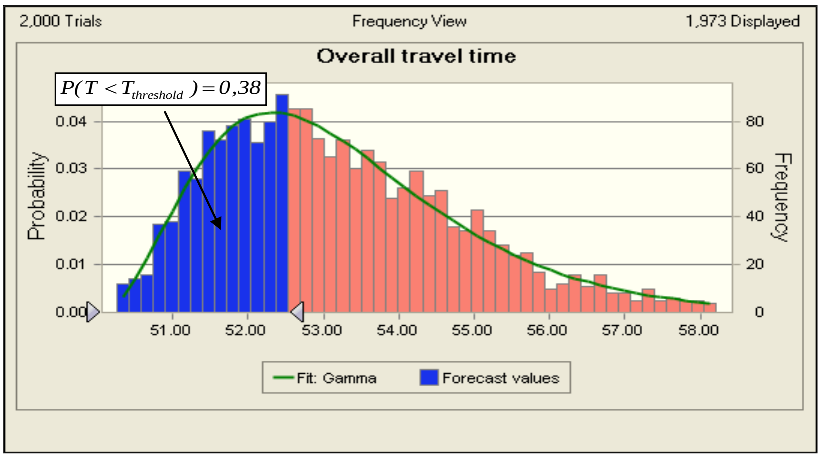

Fig.3.Forecast of overall bus travel time

great importance in respect of the optimization of planning and management of the route. For this purpose, let us suppose that the bus travel time within a specific time interval of the day follows Gamma distribution with parameters as follows: first section - $r_1 = 1,1$ and $\lambda_1 = 1$, second section - $r_2 = 2,2$ and $\lambda_2 = 1$. The simulation process for bus travel time through section 1 and section 2, and also overall bus travel time is presented on figures 1, 2 and 3. Simulation of bus travel time has two main advantages. It allows:

-determining how much the assumptions about travel time distributions (and their parameters) affect the simulation result regarding overall travel time. Figure 4 presents a sensitivity chart which shows that the second section (with assumption for Gamma distribution of bus travel time with parameters: $r_2 = 2,2$ and $\lambda_2 = 1$) has higher importance to overall travel time;

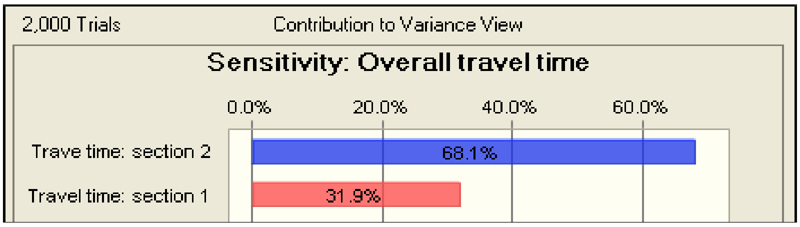

Fig.4.Contribution to variance of overall bus travel time

-determining the probability of a trip through bus route to be implemented within a designed threshold travel time (e.g. timetable travel time plus 5 % $\rightarrow T_{threshold} = 1,05.T_{timetable}$). As shown on Figure 3 this probability is $P(T < T_{threshold}) = 0,38$. In other words, the probability of delay of a bus travelling through overall route within the investigated time of day is 0,62.

3.2. Traffic safety

When decision-making is connected with traffic safety there are three key problems which should be solved, namely:

-Identifying locations with abnormal accident characteristics (locations with accident values higher than the expected values);

-Estimating the effect of one or more exposure variables (e.g. traffic and/or infrastructure characteristics) on accident occurrence;

-Identifying possible hazardous locations (requiring mitigation measures);

-Choosing between alternatives (projects) concerning safety improvement.

The identification of locations with abnormal accident characteristics could be implemented by the usage of statistical analysis of accident data regarding a group of sites with similar traffic and infrastructure characteristics (for example: geometrics, traffic volume and structure, traffic management and control, etc.). The assumption that the number of a specific type of accident at road sites (as a random variable) follows normal distribution allows us to compute the expected range of accident frequency:

$$l = \bar{x} - z\sigma \qquad (6)$$
$$u = \bar{x} + z\sigma$$

where:

$\bar{x}$ -mean (average) number of accidents estimated on the basis of statistical data for group of control sites;

z-standard normal random variable (the number of standard deviations corresponding to the required confidence level $\rightarrow z = 1{,}96$ for the *95%* confidence level);

σ-standard deviation of accident frequency estimated on the basis of statistical data for group of control sites;

u, l - upper and *lower confidence bounds* of expected range of accident frequency.

After computing the expected range of accident frequency (determined by its upper and lower bounds), the next two hypotheses regarding given location under consideration (with operational characteristics similar to those of the control sites) are may be tested:

- *Null hypothesis* (H_0) - accident rate at the location under consideration falls within confidence interval of expected range of accident frequency;
- *Alternative hypothesis* (H_1) - accident rate at the location under consideration does not fall within confidence interval of expected range of accident frequency.

Locations under study with average values of accident frequency higher than the upper confidence bound of expected value are considered as over-representing that specific type of accident. Locations with average value of accident frequency lower than the lower confidence bound of expected value are considered as under-representing that specific type of accident. Of course, over-representing locations should normally be selected for specific study and decision-making.

The effect of traffic and/or infrastructure characteristics on accident occurrence can also be assessed on the basis of accident data and by the usage of some well known statistical tools, for example statistical tests (statistical comparison): Proportionality test, H-test (Kruskal-Wallis test), t-test (Student test), Empirical Bayes method, etc. The first three tools are connected with hypothesis testing procedure which is used to make inference on special features of populations (analysis based on samples taken from these populations). Here, the basic steps of H-test will be briefly considered:

First step - Consideration of two data sets in respect of specific type of accident. Each set consists of data for accidents occurring on road location (or within road section) which is characterized with particular operational characteristics (e.g. speed limit, traffic flow structure, etc.). H-test can be used to assess whether the distribution of accidents for the first and second set are similar. In other words, the usage of this test allows assessment of the influence of one or more operational characteristics upon safety (accident occurrence).

Second step - Combination of the two data sets into one common set and ranking its elements according to their magnitude (e.g. the lowest number of accidents should be ranked with rank 1, regardless of whether it occurs on the first or second location (or road section) under study). The rest of elements are then ranked increasingly to the highest value that should have rank equal to the total number of data (for both sets).

Third step – Stating hypotheses, setting criteria for a decision, computation the test statistic and making a decision.

The null and alternative hypotheses could be stated as follows:

• *Null hypothesis* (H_0) - the probability distributions of accidents on the first and second location (or road section) are the same (there is no influence of operational characteristic under consideration upon traffic safety).

• *Alternative hypothesis* (H_1) - the probability distributions of accidents on the first and second location (or road section) are not the same (there is influence of operational characteristic under consideration upon traffic safety).

The test statistic is:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{2} \frac{R_j^2}{n_j} - 3(n+1), \qquad (7)$$

where:

n_j - total number of data measured in the *j*-th sample;

n- total sample size (total number of data in both samples);

R_j - rank sum of elements of sample *j*.

To make a decision it is necessary to compare the obtained value of test statistic *H* to the theoretical value of the χ^2 distribution (usually given in tables). The theoretical values of the χ^2 distribution depend on both the level of significance (α) and degree of freedom (*df*). The parameter *df* depends on the specific distributions that are studied: $df = k - 1$, where *k* is the number of distributions.

The null hypothesis is rejected if the obtained value *H* exceeds the theoretical χ^2 ($H > \chi^2_{\alpha, df}$). Otherwise ($H < \chi^2_{\alpha, df}$), the null hypothesis is accepted.

The identification of possible hazardous locations requiring mitigation measures and having potential for accident reduction is a very important procedure within overall road safety analysis. The usage of so called *Safety Performance Function* is one of the possible ways to fulfill this procedure [6]. This function adequately relates the expected number of accidents per unit of time and some traffic characteristics (e.g. road length, average annual daily traffic, etc.) and is developed through the use of multivariate regression techniques which incorporate the Empirical Bayesian methodology for accident data processing. The *Safety Performance Function* can be determined for fatal, injury, and property damage only accidents separately, and typically takes on the following functional form:

$$n = a_0 Q_1^{a_1} Q_2^{a_2} e^{(a_3 X_3 + a_4 X_4 + \ldots + a_k X_k)}, \qquad (8)$$

where:

n - predicted accident frequency per year at given location;

$a_0, a_1, \ldots, a_k$ - regression parameters;

Q_1, Q_2 - total average annual daily traffic on the major and minor roads, [veh/day].

Computed for a long-term period, the expected number of accidents per year (m) has the next view:

$$m = wn + (1 - w)\frac{x}{r}, \qquad (9)$$

where:

w-relative weight for the type of accident under study $\rightarrow w = \frac{k}{k + rn}$;

x-number of accidents occurred over k years at the location under study;

r-number of years within the time period under observation;

k-overdispersion parameter (related to the uncertainty in the determination of n from accident data, usually modelled by using negative binomial distribution).

The obtaining of m and n allows the determination of the safety indicator named *Potential for Safety Improvement* regarding given type of accident:

$$P_{safety\ improvement} = m - n. \tag{10}$$

The indicator $P_{safety\ improvement}$ expresses the potential for long-term average accidents to be reduced by value computed through equation (10).

Choosing between alternatives for safety improvement. This problem is connected with the determination of the economic feasibility of each set of mitigation measures and choice of the best alternative among them. A set of mathematical and statistical tools could be employed to solve this problem.

Cost-Benefit Analysis (CBA) is the most appropriate tool to support decision-making when choosing between alternatives for safety improvement and is characterized by the next key stages:

-Identification of all costs and benefits of any project for safety improvement;

-Quantification and expression of costs and benefits in monetary terms;

-Discount the future costs and benefits streams to the present value;

-Comparison of costs and benefits to determine the economic feasibility of mitigation measures.

On the basis of the linear programming (simplex method) and with the usage of CBA as analytical tool, an approach to manage the problem stated above is presented in article [2].

4. Conclusion

Only systematic, well designed and properly implemented research provides the most effective approach to the solution of many problems connected with operational reliability and safety in the field of transport. Authenticity is a key requirement for the results and conclusions following reliability (safety) research. It cannot be achieved without the usage of statistical information regarding transport accidents.

The knowledge of statistical tools and skills to collect, understand, interpret and process data regarding transport events is a fundamental requirement for experts conducting reliability and safety analysis. They have to be well prepared (theoretically educated and practically trained) to make appropriate decisions on all aspects of analysis, namely:

-Definition of the problem under consideration and results that have to be obtained;

-Identification of the ways (approaches and methods) to fulfil the analysis;

-Identification of the need for statistical information;

-Choice of the most appropriate statistical method to achieve the results stated initially;

-Implementation of the sequence of procedures within the method chosen.

This paper discusses some fundamental statistical approaches which could be successfully used to solve specific operational problems.

REFERENCES

[1] Fan, Y., Kalaba, R., Moore, J. *Arriving on time*. Journal of Optimization Theory and Applications, 127 (3), 2005.

[2] Georgiev, N. *Decision-making information concerning transport safety improvement in conditions of risk and uncertainty*. Thirteenth International Scientific Conference "CRISES SITUATIONS SOLUTION IN SPECIFIC ENVIRONMENT", Žilina, Slovakia, 2008.

[3] Hauer, E. *Observational Before–After Studies in Road Safety, Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*, Pergamon 1997.

[4] Lomax, T., Turner, S., Schrank, D. *Selecting Travel Reliability Measures*. Texas Transportation Institute and Cambridge Systematics Inc., 2003.

[5] Nie, Y., Fan, Y. *Optimal Routing for Maximizing the Travel Time Reliability*. Springer Science and Business Media, LLC 2006.

[6] Paniati, J., True, J. *Interactive Highway Safety Design Model*. Transportation Research Board, (National Research Council), Washington, 1996.

Článok recenzovali dvaja nezávislí recenzenti.