Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Random delays forming in the dense train flow

Boris Davydov[a], Vladimir Chebotarev[b], Kseniya Kablukova[b,]*

[a] *Far Eastern State Transport University, Seryshev st. 47, 680021, Khabarovsk, Russia*
[b] *Computing Center, Far Eastern Branch of RAS, Kim Yu Chen st. 65, 680000, Khabarovsk, Russia*

Abstract

One of the main characteristics of the punctuality of railway traffic is the deviation of actual arrival time from the planned one. The reasons for occurrence of the arrival deviations are departure deviations and scattering of travel times over the route. Two stochastic models of arrival deviations formation are proposed in this article. In the framework of these models, we obtain the formulas for finding the probabilistic distribution of examined deviations. A probabilistic interpretation of the experimental data, received from Russian Railways, on the trains' movement is given. Empirical data are compared with the conclusions of the proposed theoretical models.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: train traffic; stochastic model; optimal rescheduling; train delay propagation; probabilistic modeling; operative management.

1. Introduction

Deviations from the schedule and inter-arrival time are important flow characteristics which affect level of service and capacity of the transportation system. Traffic situations which are hard to observe in the real world can be investigated with the use of mathematical simulation. The analytical method for determining the probability distribution of the arrival times, for instance, density function (pdf), assumes the existence of several random variables whose influence is summed. Such a mechanism arises from the concept that a vehicle run is a chain of operations each with their own random realization time. Further, digital computer simulation can be used to analyze the train traffic and identify the optimal adjustments.

The stochastic model takes into account random processes the train traffic on open tracks and at stations located on the managing area. In addition, deviations from the schedule at the input boundary of the section are taken into account when delays analyzing. In most published papers, the processes of forming the delay distributions at stations

* Corresponding author. Tel.: +7-914-181-6301.
E-mail address: kseniya0407@mail.ru

2352-1465 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

and on interstations are investigated separately. In this case it does not taking into account such a general that join these processes and provides a unified approach to calculate the arrival distribution. Numerous studies show the scattering of the duration the primary delays obey an exponential law. As for the travel times, here the distribution behavior is more complicated. This is due to the change in the conditions of the train passage which is of a random character. In addition, there is a component of on-line adjustments carried out by the driver and the train manager. We guess the result of this joint influence is that distribution of the arrival time at the station in most cases different from the exponential one.

The station also has a stabilizing mechanism that aims to reduce the delays. The result is a decrease in the scattering width of the departure times.

Knowledge of random deviations formation mechanism makes it possible to solve the following problem: to find the distribution of the arrival time. Its solution allows to reduce the computational effort in solving the problem for a large-scale railway network.

In this paper, we propose two models for the formation of arrival time deviations in the presence of two perturbing factors, namely, the scattering of the departure times and the random travel time along the track. After the literature review, which is given in Section 2, we describe a process of formation the arrival times scattering. The process is caused by operational adjustments performed by the driver. A model of the output distribution is constructed in Section 3. It takes into account the random nature of departure and travel times.

The results of the simulation are consistent with the following property of any particular operation on open paths: the distribution of the deviation of the real time of the operation from the planned one (a) is exponential if one dominant factor is present and (b) is smoother than the exponential, in the sense that the density function, except of the right, has a left branch, if several independent factors of a different nature present too.

Disturbances of the second kind include delays caused by technical malfunctions and weather conditions changing. Abovementioned regularities can be observed for the flow of any vehicles, in particular, for the movement of cars on freeways. The statistical data, which are given in Section 4, are consistent with these theoretical conclusions and heuristic reasoning.

2. Literature review

Railway traffic usually is represented by a timed event graph that allows computing delay propagation in large network in a short time (Cacchiani et al. (2014)). Rescheduling model created by this methodology unable to take a random change in the current condition of infrastructure and weather into account. Stochastic model allows to eliminate this drawback by use the scattering of trajectory element duration to arrival time prediction. The probability density function (pdf) of arrival times at a stations or operating points is calculated using analytical models or simulations. In recent years, the uncertainty of train event times has been recognized as one of the major obstacles for computing feasible and implementable solutions for rescheduling problem (Corman and Meng (2014)).

One of the first papers considering this problem (Muhlhans (1990)) proposed analytical method for the delay estimating by a convolution of pdf the initial (inlet area) deviations. The analysis is limited because it used the assumption of uniform distribution the secondary delays. In reality, this assumption is not confirmed. Most stochastic models to analyze propagation of delays focused on single track routes or on simple junctions.

Classical stochastic models for the propagation of delays had been studied intensively, most importantly in Carey and Kwieciński (1994, 1995). The papers proposed the approximation of delay distributions to reduce the computational effort and studied the error propagation for such approximations. Used model considers the total running time as sum of the partial intervals of the moving trains. Each interval is considered as a random variable with the exponential distribution. The studies do not consider the pdf of the headways at a section exit and the mechanism of the delay propagation through the chain of trains.

The paper Meester and Muns (2007) represents trains traffic in a form occupying an intermediate position between the macro-models and models with very detailed description of the process, i.e., micro-models. Approximation method is proposed for the exact representation of delay distributions. Cumulative distribution is calculated from the sequence of activities, which is determined by a stochastic event-graph.

A stochastic analytical model based on a double track line was proposed in Huisman and Boucherie (2001). The paper investigates delays to a fast train caught behind slower ones by capturing both scheduled and unscheduled movements. This is modeled as an infinite server. The running time distributions for each train service are obtained by solving a system of linear differential equations by assuming deterministic free running times. This model give a good insight into delay propagation on one line or a simple network, but become too complicated to handle when dealing with large scale real-world networks. The approach for predicting waiting times using queuing system with a semi-Markovian kernel was presented in Wendler (2007). Description of the service process is based on an application of the theory of blocking times and minimum headway times.

A stochastic model for delay propagation and forecasts of arrival and departure events which is applicable to all kind of public transport was proposed in Berger et al. (2011). The model includes the general train waiting policies on stations and considers discrete distributions of travel time profiles which depend on the departure time. Use of this model allows you to define the order of departure of the train and the distribution of the respective moments of departure. The extension of this study is made in the paper Oblakova et al. (2016) which describes the analytical method for calculating the pdf of delays based on vehicle chain analysis.

Discrete stochastic model of disturbed train traffic was also used in Keyhani et al. (2012) which enhances the formulas of the above mentioned paper to calculate probability distributions for train connections. Another model proposed in Yuan (2006) is also designed to assess the knock-on delays. The analytical probabilistic model takes into account the duration of stops, stochastic relationship between the trajectories of trains, speed fluctuations and dynamic propagation of delays. New scheme proposed for the exact determination of parameters the theoretical distribution model based on the maximum likelihood method.

A number of studies considered the model that comprehends two train runs, each with a stopping at a station and with transferring passengers (Goverde (2010)). The model illustrated as a Petri net graph. The approach proposed in Büker and Seybold (2012) uses a probabilistic operational graph, which considers transfer operation and conflict situations associated. The authors argue a mesoscopic modelling of traffic is the purposeful approach to compute the delay propagation. Events that in reality lead to changes in the delay such as interlinking conflict are mapped in the model by manipulating the pdf of elementary activities. Using of both conditional and unconditional convolution and of «excess beyond» operations is specifically made. Various types of delay modeled by means of cumulative distribution function. A model created by Büker and Seybold (2012) is logically incomplete. The authors consider only station and crossing as the check points where conflicts can occur. To complete the model you must use a resolution of the sequence-of-trains conflicts which appear at open tracks.

A significant number of researchers represent the train run as a Markov chain with state transitions in discrete moments associated with arrivals and departures at the scheduled stop. After every registered departure or arrival event, the conditional probability distributions are updated with respect to the essential assumption for Markov process.

Various models which describe the functions of the frequency and length of nonscheduled delays are discussed in prior works. Thus, in the papers (Schwanhäußer (1974), Wendler (2007), Goverde (2005), Kecman et al. (2015)) the exponential probability distribution is used as a stochastic model of the train movement which is violated when additional reasons occur. Obviously, accumulation of the individual random variables leads to non-exponential behavior of the arrival time pdf. This problem is explored in Chebotarev et al. (2015) and requires further study especially for the mixed flow of passenger and freight trains.

There are very few studies that analyze the effect of short-term disturbances on train traffic and the personnel activity to compensate for accidental impacts. Most of the existing approaches assume fixed probability distributions for the estimation of process times and do not consider the effect that real-time information on train positions and delays may have on (the parameters of) the corresponding distributions. In order to create realistic online tools for real-time traffic management, the dynamics of uncertainty of delays needs to be considered. The first stochastic approach that to some extent exploits the current traffic conditions to reduce uncertainty is presented by Berger et al. (2011). The authors proposed several theoretical probability distributions of running times, depending on departure time for each train type. An approach that considers the dynamics of uncertainty of train delays was presented by Bauer and Schobel (2014). The authors developed a «delay generator» for the purpose of integrating uncertainty in online traffic management. But they used a static pdf representation that describes a delay no corresponding to reality in its form. The paper Davydov et al. (2017) summarizes the work that is carried out in this important direction. The results of further study of this problem are proposed in this paper.

3. Two models of arrival delay forming

The first model. Let us formulate the main assumptions in our first two-train model:

a) the actual departure times of trains coincide with scheduled ones,

b) for each train the actual travel time between stations is a random variable,

c) the minimum permissible distance between trains is equal to some fixed $s_0 > 0$; as soon as the distance between the trains becomes equal to s_0, the speed of train $i+1$ decreases and becomes equal to the speed of train i.

We shall use the following notations. For every train i:

$d_{S_1}^{(i)}$ is the planned departure time from station S_1,

$a_{S_2}^{(i)}$ is the planned arrival time at station S_2,

$a_{S_2}^{(i)} + \xi_i$ is the actual arrival time at station S_2,

$\rho_{S_1S_2}^{(i)}$ is the actual travel time from station S_1 to station S_2.

The distribution functions (cumulative distribution functions) of the random variables ξ_i and $\rho_{S_1S_2}^{(i)}$ are denoted by $V_i(t)$ and $R_{S_1S_2}^{(i)}(t)$ respectively, i.e. $V_i(t) = \mathrm{P}(\xi_i < t)$, $R_{S_1S_2}^{(i)}(t) = \mathrm{P}\left(\rho_{S_1S_2}^{(i)} < t\right)$ (by definition, we assume that the distribution functions are continuous from the left). Note that $\mathrm{P}\left(\rho_{S_1S_2}^{(i)} > 0\right) = 1$, and $\xi_i \in R \equiv (-\infty, \infty)$.

Let trains i and $i+1$ move from station S_1 to station S_2. We will distinguish between the following two cases. The first one is that during the entire movement the distance between the trains is strictly greater s_0, and the second, that at some moment of time the distance will coincide with s_0.

Obviously, in the first case train i does not affect the movement of the train $i+1$. Put, for instance, $i=1$. In Fig. 1 the bundles of straight lines emanating from the points $d_{S_1}^{(1)}$ and $d_{S_1}^{(2)}$ show a scattering the travel times of trains 1 and 2 between stations S_1 and S_2. And the bold points on the line S_2 denote possible values of the trains' arrival times. Since the current distance between the trains (we denote it by r) is greater s_0 at each moment $t \in \left[d_{S_1}^{(2)}, a_{S_2}^{(1)} + \xi_1\right]$, the second train moves unhindered along the original trajectory to the end of the route. In this case, the actual arrival time $a_{S_2}^{(2)} + \xi_2$ of train 2 to station S_2 is equal to $d_{S_1}^{(2)} + \rho_{S_1S_2}^{(2)}$ (see Fig. 1). Consequently, $\xi_2 = d_{S_1}^{(2)} - a_{S_2}^{(2)} + \rho_{S_1S_2}^{(2)}$, and then

$$V_2(t) = \mathrm{P}\left(d_{S_1}^{(2)} - a_{S_2}^{(2)} + \rho_{S_1S_2}^{(2)} < t\right) = R_{S_1S_2}^{(2)}\left(t + a_{S_2}^{(2)} - d_{S_1}^{(2)}\right). \quad (1)$$

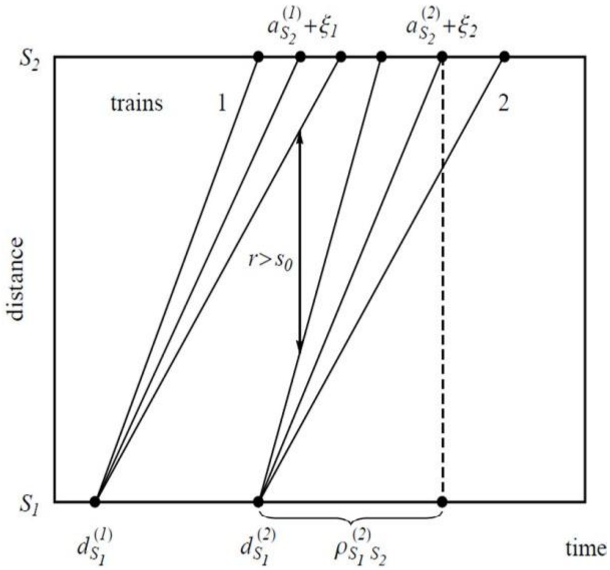

Fig. 1. Case 1: there are no an influence of the first train to the second one

In the second case the second train is forced to decrease its speed and equalize it with the speed of the first train. In Fig. 2, the dashed line continuations show the possible trajectories of the second train in the case, as if the influence of the first train would be absent. If, at some point of time $t \in \left[d_{S_1}^{(2)}, a_{S_2}^{(1)} + \xi_1 \right]$, the distance r is equal to s_0, the second train decreases its speed to the speed of first train.

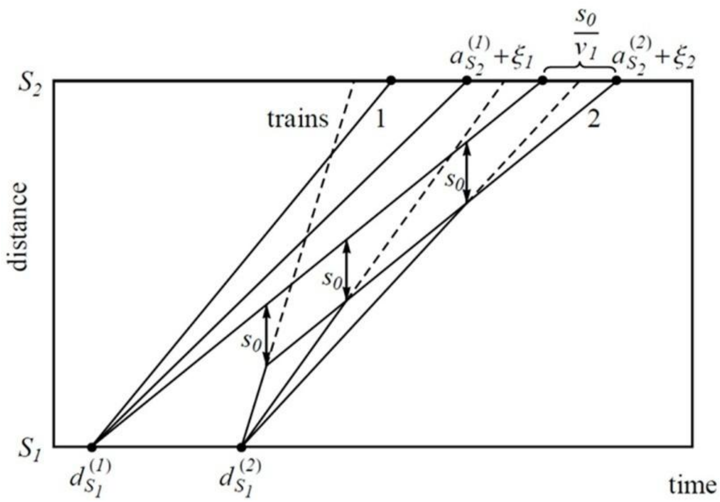

Fig. 2. Case 2: the second train decreases its speed to the speed of the first train

Now, in addition to the main assumptions a) – c), we assume that

d) the speed of train i is equal to $v_i := \dfrac{S_2 - S_1}{\rho_{S_1 S_2}^{(i)}}$.

Since $a_{S_2}^{(2)}+\xi_2=a_{S_2}^{(1)}+\xi_1+\frac{s_0}{v_1}$ (see Fig. 2), and $a_{S_2}^{(1)}+\xi_1=d_{S_1}^{(1)}+\rho_{S_1S_2}^{(1)}$, then $a_{S_2}^{(2)}+\xi_2=d_{S_1}^{(1)}+\rho_{S_1S_2}^{(1)}+\frac{s_0}{S_2-S_1}\rho_{S_1S_2}^{(1)}$. Consequently, $\xi_2=d_{S_1}^{(1)}-a_{S_2}^{(2)}+\frac{S_2-S_1+s_0}{S_2-S_1}\rho_{S_1S_2}^{(1)}$. Thus,

$$V_2(t)=R_{S_1S_2}^{(1)}(u),\quad \text{where}\quad u=\frac{(S_2-S_1)\left(t+a_{S_2}^{(2)}-d_{S_1}^{(1)}\right)}{S_2-S_1+s_0}. \tag{2}$$

The second model. This model differs from the first one by the following generalization of the assumption a): we suppose that the actual departure times from station S_1 do not necessarily coincide with the planned ones, i.e. the i-th train departs at the moment $d_{S_1}^{(i)}+\delta_i$, where δ_i is the random variable. Now in the first case, instead of $\xi_2=d_{S_1}^{(2)}-a_{S_2}^{(2)}+\rho_{S_1S_2}^{(2)}$, we have $\xi_2=d_{S_1}^{(2)}+\delta_2-a_{S_2}^{(2)}+\rho_{S_1S_2}^{(2)}$, and in the second one, instead of $\xi_2=d_{S_1}^{(1)}-a_{S_2}^{(2)}+\frac{S_2-S_1+s_0}{S_2-S_1}\rho_{S_1S_2}^{(1)}$, we have $\xi_2=d_{S_1}^{(1)}+\delta_1-a_{S_2}^{(2)}+\frac{S_2-S_1+s_0}{S_2-S_1}\rho_{S_1S_2}^{(1)}$.

In addition, we introduce the following technical assumption:

e) the random variables δ_i and $\rho_{S_1S_2}^{(i)}$ are independent.

Since the distribution of independent random variables is the convolution of initial distributions, this assumption leads to the following formulas. In the first case we have

$$V_2(t)=\mathrm{P}\left(d_{S_1}^{(2)}-a_{S_2}^{(2)}+\delta_2+\rho_{S_1S_2}^{(2)}<t\right)=\int_0^\infty \mathrm{P}\left(\delta_2<t-d_{S_1}^{(2)}+a_{S_2}^{(2)}-x\right)dR_{S_1S_2}^{(2)}(x), \tag{3}$$

and in the second one,

$$V_2(t)=\mathrm{P}\left(d_{S_1}^{(1)}-a_{S_2}^{(2)}+\delta_1+\rho_{S_1S_2}^{(1)}\frac{S_2-S_1+s_0}{S_2-S_1}<t\right)=\int_0^\infty \mathrm{P}\left(\delta_1<t-d_{S_1}^{(1)}+a_{S_2}^{(2)}-x\frac{S_2-S_1+s_0}{S_2-S_1}\right)dR_{S_1S_2}^{(1)}(x). \tag{4}$$

We shall call the distributions of the random variables δ_i and $\rho_{S_1S_2}^{(i)}$ by input ones, and the distribution of ξ_i by output ones, i = 1, 2.

Example. Let δ_i, i = 1, 2 have one and the same shifted exponential distribution with parameters $\lambda_1>0$ and $b_1\in R$, i.e.

$$\Theta(x)\equiv \mathrm{P}(\delta_i<x)=I(x>b_1)\left(1-\exp\{-\lambda_1(x-b_1)\}\right), \tag{5}$$

and $\rho_{S_1S_2}^{(i)}$, i = 1, 2 have one and the same gamma distribution with parameters $\alpha>0$ and $\beta>0$, i.e.

$$R_{S_1S_2}^{(i)}(x)=I(x>0)\frac{\gamma(\alpha,x/\beta)}{\Gamma(\alpha)}, \tag{6}$$

where $\gamma(a,z)=\int_0^z e^{-t}t^{a-1}dt$ is lower incomplete gamma function. The choice of distributions (5) and (6) is due to statistical data. As initial parameters we take the following values, calculated from the statistical sample:

$$\lambda_1 = 0.86\ [1/\text{min}],\quad b_1 = -0.2\ [\text{min}],\quad \alpha = 41.51,\quad \beta = 0.28\ [\text{min}]. \tag{7}$$

It should be noted that we do not know which of the distribution, (3) or (4), agrees better with the real output distribution at a given track. Therefore, we check both cases.

First, let us set the constants which are determined by the train schedule and the distance between the stations:

$$d_{S_1}^{(1)} = 6.25\ [\text{h}],\quad d_{S_1}^{(2)} = 6.32\ [\text{h}],\quad a_{S_2}^{(2)} = 6.48\ [\text{h}],\quad S_2 - S_1 = 13\ [\text{km}],\quad s_0 = 4.67\ [\text{km}]. \tag{8}$$

In Fig. 3a the graphs of distribution functions $V_2(t)$ from (3) (curve 1), $\Theta(x)$ (curve 2) and $R_{S_1S_2}^{(2)}(x)$ (curve 3) are depicted. Fig. 3b shows pdf of examined random variables.

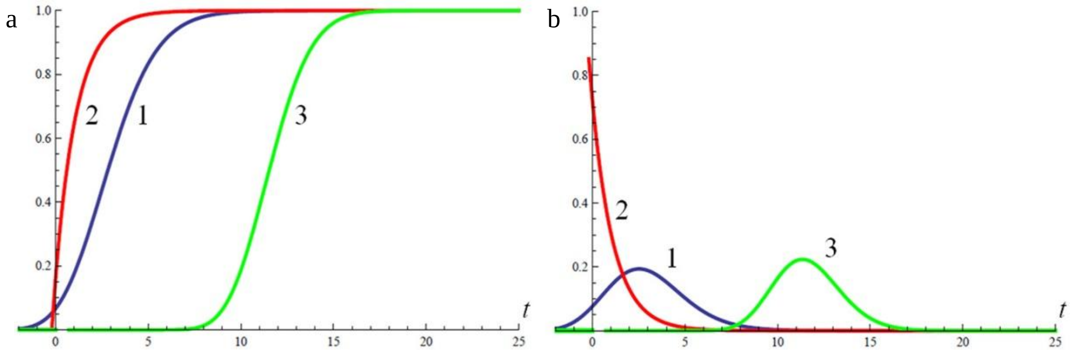

Fig. 3. Input and output distribution functions (a) and pdf (b) in the case when the second train does not change its speed

Find out the mean and variance of the output distribution (curves 1 in Fig. 3). Taking into account properties of exponent and gamma distributions we have $\mathrm{E}\delta_i = \frac{1}{\lambda_1} + b_1$, $\mathrm{D}\delta_i = \frac{1}{\lambda_1^2}$, $\mathrm{E}\rho_{S_1S_2}^{(2)} = \alpha\beta$, $\mathrm{D}\rho_{S_1S_2}^{(2)} = \alpha\beta^2$. First, it follows from here that $\mathrm{E}\xi_2 = d_{S_1}^{(2)} + \mathrm{E}\delta_2 - a_{S_2}^{(2)} + \mathrm{E}\rho_{S_1S_2}^{(2)} = d_{S_1}^{(2)} + \lambda_1^{-1} + b_1 - a_{S_2}^{(2)} + \alpha\beta$. Moreover, by independence of δ_2 and $\rho_{S_1S_2}^{(2)}$, we have $\mathrm{D}\xi_2 = \mathrm{D}\delta_2 + \mathrm{D}\rho_{S_1S_2}^{(2)}$. Therefore,

$$\mathrm{E}\xi_2 = 6.32[\text{h}] + \frac{1}{0.86}[\text{min}] - 0.2[\text{min}] - 6.48[\text{h}] + 41.51\cdot 0.28[\text{min}] \approx 2.99[\text{min}], \tag{9}$$

$$\mathrm{D}\xi_2 = \frac{1}{0.86^2}\left[\text{min}^2\right] + 41.51\cdot 0.28^2\left[\text{min}^2\right] \approx 4.61\left[\text{min}^2\right]. \tag{10}$$

In Fig. 4a the graphs of distribution functions $V_2(t)$ from (4) (curve 1), $\Theta(x)$ (curve 2) and $R_{S_1S_2}^{(1)}(x)$ (curve 3) are depicted. The graphs of corresponding density functions are depicted in Fig. 4b.

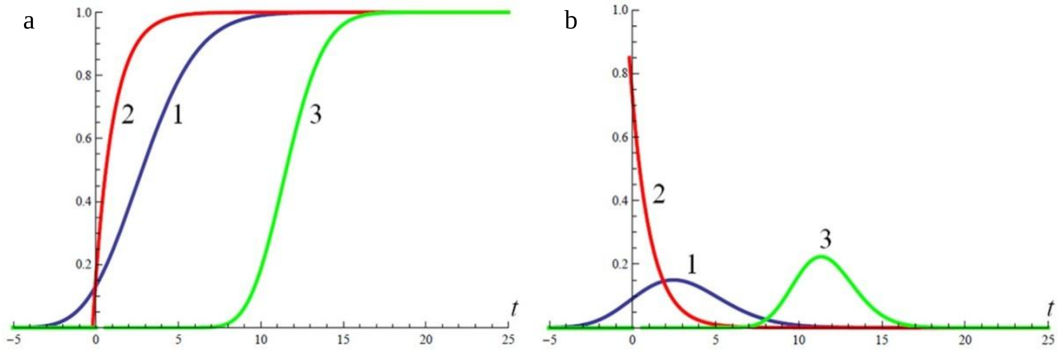

Fig. 4. Input and output distribution functions (a) and densities (b) in the case when the second train changes its speed by v_1

Now we suppose, $\xi_2 = d_{S_1}^{(1)} + \delta_1 - a_{S_2}^{(2)} + \frac{S_2 - S_1 + s_0}{S_2 - S_1} \rho_{S_1 S_2}^{(1)}$. Taking this into account and using properties of mathematical expectation and variance we obtain the mean and variance of the output distribution (curves 1 in Fig. 4):

$$\mathrm{E}\xi_2 = 6.25[\mathrm{h}] + \frac{1}{0.86}[\min] - 0.2[\min] - 6.48[\mathrm{h}] + \frac{13+4.67}{13} 41.51 \cdot 0.28[\min] \approx 2.96[\min], \tag{11}$$

$$\mathrm{D}\xi_2 = \frac{1}{0.86^2}\left[\min^2\right] + \left(\frac{17.67}{13}\right)^2 41.51 \cdot 0.28^2 \left[\min^2\right] \approx 7.36\left[\min^2\right]. \tag{12}$$

According to the sample from the distribution of random variable ξ_2, the sample mean is equal to 2.5 min, and the sample variance is equal to $4.18\,\min^2$. If we compare these statistical characteristics with the values calculated on the basis of the model (in case 1 these are the values (9) and (10), in case 2 the values (11) and (12)), we can observe that the statistical characteristics are closer to the values of (9) and (10). Therefore, one can conclude that in the example under consideration the preceding train likely does not affect the movement of the next one (case 1), and arrival delays occur due to other factors.

4. Experimental results

In some previous papers (see as example, Kecman et al. (2015), Yuan (2008)), histograms of arrival and departure times derived from statistical data are given. The study of these histograms shows that the vast majority of considered events occur before their scheduled time. It has been shown that the histograms of event times and process times are often skewed to the right that is, they have an intense right "tail".

Studying of these histograms shows that arrival event often occurs before the time provided by the schedule. Checking the statistical hypothesis about the distribution of the random arrival deviation ξ shows that the type of its probability density function is different from the exponential one. Although the asymmetry coefficient, as well as an exponential distribution's one, is positive, a probability density function of the random variable ξ has a left branch. Note that the left branch is often shorter than the right one. A similar kind of distribution is also inherent to the train travel time by the route.

It is shown in Davydov et al. (2016) that the histogram of arrival deviations is well approximated by the density function of the asymmetric distribution with a positive asymmetry coefficient, but with the left branch. Perhaps the reason for this is the combined effect of a multitude of random perturbations arising on open tracks.

At the same time, the probabilistic distribution of departure deviations is often exponential (generally, with a shift). This fact is confirmed by statistical analysis. It can be assumed that this fact is a consequence of a strict control over the departure times by dispatchers.

The train manager and the driver are active to ensure the most accurate adherence the departure timetable. The precision management of passengers boarding / alighting is one of the key mechanisms that ensure punctuality (Yuan (2008)).

As a result of our experimental studies, it can be concluded that the on-line adjusting affect the type of the distribution of the arrival deviation from the planned one in a similar way.

We have investigated in detail the behavior of probability distributions of the departure and arrival times of suburban trains, as well as the travel time on the Moscow-Tver line. This line has a length of 167 km. The intensity of traffic reaches fifteen trains at every rush hour.

In addition, we have analyzed the statistical characteristics of mixed train traffic on some section of the Trans-Siberian mainline of the length about 400 km. Heavy freight trains make the main part of the services passing on this line. Container trains also move along the mainline according to the strict timetable. The study showed that the regularities in the formation of deviations from the schedule are observed for passenger and freight trains**.**

As the main regularity, it is necessary to note the tendency to increase deviations from the schedule when the train moves along the route. It is connected with the accumulation of small random deviations. Such behavior of the average arrival deviations is observed in those tracks where the adjusting activity of the drivers and dispatchers is relatively weak. As a rule, this ivent happens on the initial section of the route (see Fig. 5). Here you can see that the average delays increase at the stations. This is due to the process of boarding / alighting passengers or accidental deviations in the technology of handling freight trains. An opposite behavior of delays is observed in open tracks. Generally, the efforts of driver cause compensation of deviations at open tracks.

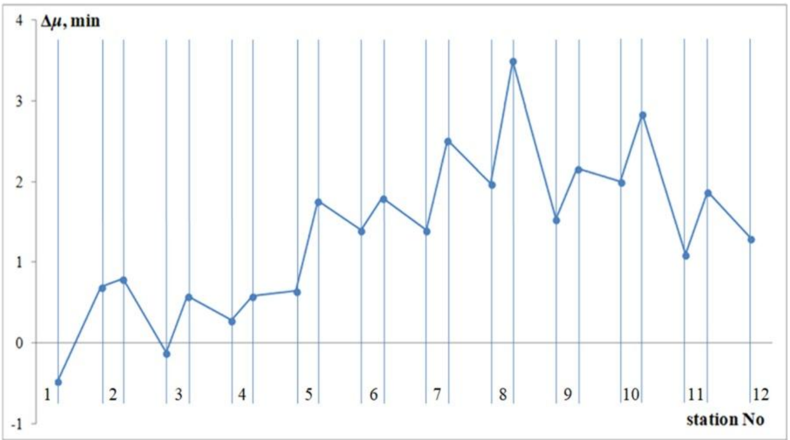

Fig. 5. Behavior of the mean values of departure and arrival deviations along the route for the suburban line

The accuracy of arrival at the terminal station is one of the important indicators of the quality of train traffic. Therefore, the staff makes extra efforts to ensure that the trains arrive on time. This is reflected in the fact that the mean of lateness decreases at the last sections of the route. In addition, trains frequently depart before the scheduled times in order to have a time supplements for compensation random delays as it is reflected in the Fig. 5.

The study of statistical data allows us to assume that drivers and dispatcher are not able to eliminate random disturbances, which causes a continuous increase in standard deviation (SD) of arrival times (see Fig. 6).

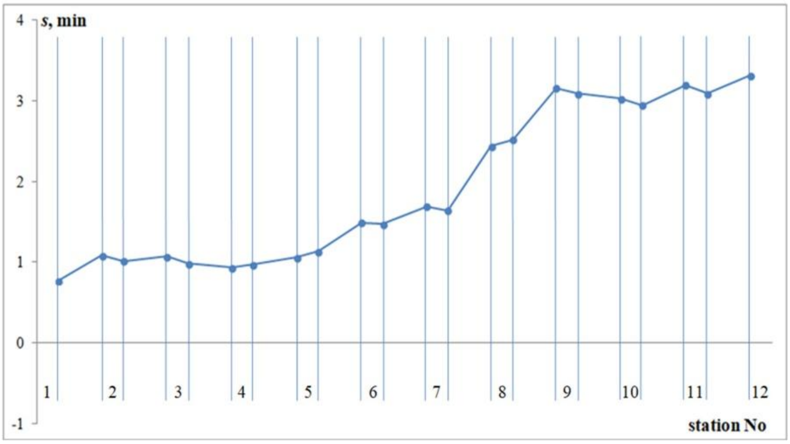

Fig. 6. Behavior of SD of departure and arrival deviations along the route for the suburban line

This pattern is typical for the movement of both passenger and freight services. The behavior of deviations from the schedule when the freight trains passing along the Trans-Siberian are similar to that occurs in passenger train traffic (see Fig. 7).

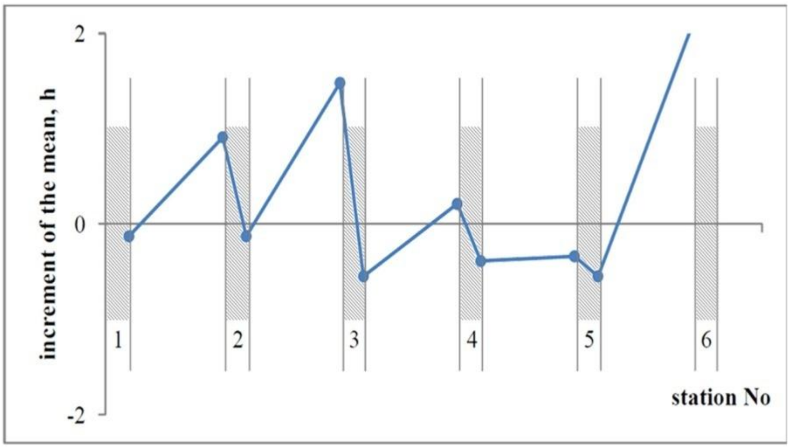

Fig. 7. Behavior of mean value of the arrival deviations at open tracks and stations for heavy haul trains

We carried out statistical analysis of real deviations in a train movement. It allows us to assume that the deviation of the actual completion time of any operation from scheduled has exponential distribution in those cases when the situation is under the staff control and there is no set of factors of equal strength of action. In particular, such distribution is observed when train arrives at passenger terminals during the day time when there is no intensive flow of passengers and the corresponding disturbances at the stations. For example, deviations of arrival times are distributed according to an exponential law on four of the thirteen Moscow-Tver lines' stations. This is observed in the evening hours at peripheral stations, where the flow of passengers is small.

In peak periods, the form of the output distribution changes in the following way: a left branch of the probability density appears (see Fig. 8).

In rush hours such type of distribution is observed at almost all stations. Actual behavior of the arrival deviations scattering corresponds to the theoretical result, which was obtained earlier in Section 3.

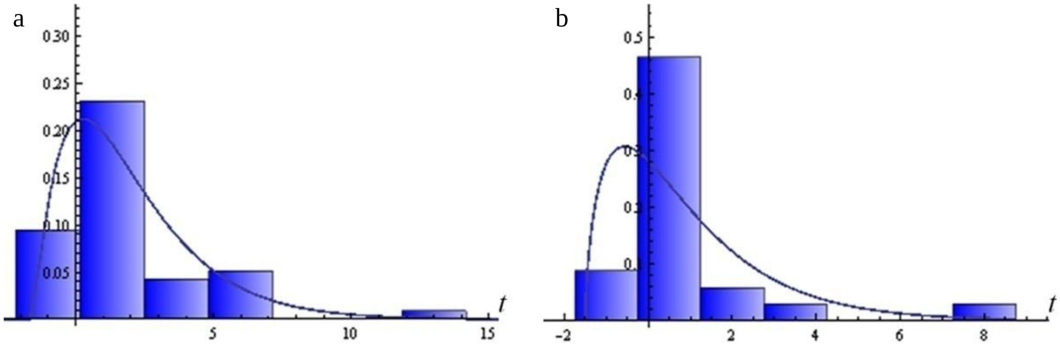

Fig. 8. Line Tver-Moscow, histograms for arrival deviations at station 6 for (a) morning peak; (b) evening peak

Fig. 9 depicts the deviation histograms for a) the departure times of freight trains from a large station and b) the arrival times at the destination. Train movement occurs in the night period, when there are no breaks for repair work. Observations show that train departure time from the cargo terminal is significantly dispersed due to the complex process of forming the composition of the carriages. Therefore, one can conclude: the distribution of the departure time deviation is the result of the impact of a variety of random independent factors. Further, when a train moves along the railway line, other random independent factors influence its movement characteristics. That is why a similar conclusion can be made for the distribution of the arrival time deviations. The histograms in Fig. 9 are statistical analogs of smoother distributions than an exponential distribution, which is natural given the central limit theorem. Note that the densities in cases a) and b) can be either with positive asymmetry or with negative asymmetry.

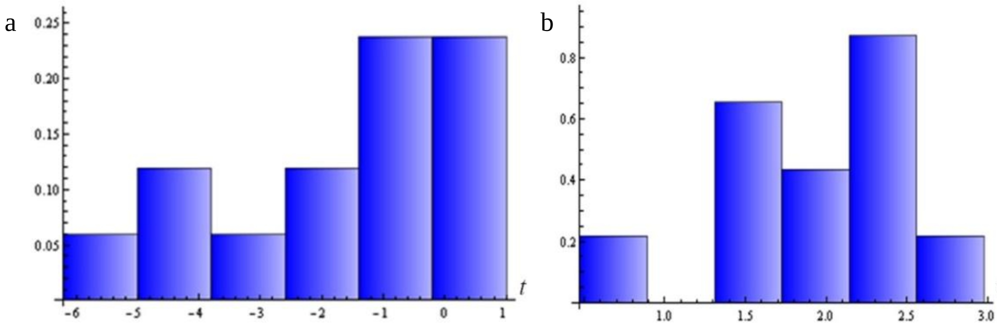

Fig. 9. Line Khabarovsk-Ruzhino (Far East), histograms for (a) departure deviations; (b) arrival deviations

According to Fig. 9 there are significant deviations of actual departure and arrival times from scheduled ones. Fig. 9a illustrates the case of early departures: a plenty of negative departure deviations is observed. The mean of departure deviations is equal to -1.87 [h]. Despite early departures, almost all arrivals are delayed (see Fig. 9b): i.e. positive arrival deviations are observed. The mean of arrival deviations is equal to 1.96 [h]. If the departures were on schedule or with delays (not early ones), the arrival times would have moved further to the right along the time axis, the duration of the arrival deviations would have increased. Therefore, in order to compensate the large arrival delays, that arise due to a variety of random factors, the dispatchers decide on early departures of trains.

5. Conclusions and future work

Stochastic modeling of railway personnel activity show, the conventional adjusting measures during the running phase lead to keeping of the unchangeable character of arrival distribution, most often the exponential one.

We assume, existence of single dominant process leads to a simpler distribution of the arrival of times such as exponential one. Adding other random processes leads to a change in the form of the distribution and to an increase in its dispersion.

The analysis shows that gamma distribution is suitable approximation for the arrival time scattering. This type of distribution arises when there are several random interacting processes that are caused by disturbances.

These conclusions are confirmed both by theoretical constructions and by results of the real statistics investigation.

Section 3 shows that the density of arrival deviations can take the form of a single-vertex density with the left and right branches (relative to the mode). This is consistent with the type of histogram of the output deviations from the schedule (Section 4).

The regularities of formation the arrival time distribution are common for the passenger and freight train traffic.

We suppose that the described mechanism for the delay formation at open tracks is analogous to that one is existed at stations. The objective of our subsequent research is to model the relevant situation when there are a lot of random disturbing factors at the railway station.

Acknowledgements

The research is partially funded by the JSC Russian Railways (grant 2016 for development of the scientific school) and by the Far Eastern State Transportation University.

References

Bauer, R., Schobel, A., 2014. Rules of thumb: practical online-strategies for delay management. Public Transport 6, 1-2, 85–105.

Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M., 2011. Stochastic Delay Prediction in Large Train Networks. In: Caprara, A., Kontogiannis, S. (Eds.), 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'11), 100–111.

Büker, T., Seybold, B., 2012. Stochastic modelling of delay propagation in large networks. Journal of Rail Transport Planning & Management, 2, 34–50.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. Overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research, Part B, 63, 15–37.

Carey, M., Kwieciński, A., 1994. Stochastic approximation to the effects of headways in knock-on delays of trains. Transportation Research, Part B, 28, 251–267.

Carey, M., Kwieciński, A., 1995. Properties of expected costs and performance measures in stochastic models of scheduled transport. European Journal of Operational Research, 83, 182–199.

Chebotarev, V., Davydov, B., Dynkin, B., Kablukova, K., 2015. Prediction of the train traffic when random failures occur, 6th International Conference on Railway Operations Modelling and Analysis, RailTokyo 2015. Tokyo, Japan, paper #043.

Corman, F., Meng, L., 2014. A review of online dynamic models and algorithms for railway traffic rescheduling. Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems, 99, 1–11.

Davydov, B.I., Chebotarev, V.I., Kablukova, K.S., 2016. Probabilistic model of the vehicle movement. Transport: Science, Technology, Management, 11, 13–19 [Russian].

Davydov, B.I., Chebotarev, V.I., Kablukova, K.S., 2017. Stochastic model for the real-time train rescheduling. International Journal of Transport Development and Integration, 1, 3, 307–317.

Goverde, R.P.M., 2005. Punctuality of Railway Operations and Timetable Stability Analysis. TRAIL Thesis Series no. T2005/10, The Netherlands TRAIL Research School, 310 p.

Goverde, R.P.M., 2010. A delay propagation algorithm for large-scale railway traffic networks. Transportation Research, Part C, 18, 3, 269–287.

Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. Transportation Research, Part B, 35, 271–292.

Kecman, P., Corman, F., Meng, L., 2015. Train delay evolution as a stochastic process, 6th International Conference on Railway Operations Modelling and Analysis, RailTokyo 2015. Tokyo, Japan, paper #007.

Keyhani, M.H., Schnee, M., Weihe, K., Zorn, H-P., 2012. Reliability and Delay Distributions of Train Connections. In: Delling, D., Liberti, L. (Eds.), 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'12), 35–46.

May, A.D., 1990. Traffic Flow Fundamentals. Prentice Hall Inc., 11–51.

Meester, L.E., Muns, S., 2007. Stochastic delay propagation in railway networks and phase-type distributions. Transportation Research, Part B, 41, 218–230.

Muhlhans, E., 1990. Berechnung der Verspatungsentwicklung bei Zugfahrten. Eisenbahntechn, Rundschau ETR, 39, 7/8, 465–468.

Oblakova, A., Al Hanbali, A., Boucherie, R.J., van Ommeren, J.C.W., Zijm, W.H.M., 2016. Exact expected delay and distribution for the fixed-cycle traffic-light model and similar systems in explicit form. Technical Report, September 2016, 34 p. DOI: 10.13140/RG.2.2.13001.75360.

Schwanhäußer, W., 1974. Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn, Publications by the Transport Science Institute at RWTH Aachen, 20.

Wendler, E., 2007. The scheduled waiting time on railway lines. Transportation Research, Part B, 41, 2, 148–158.

Yuan, J., 2006. Stochastic Modelling of Train Delays and Delay Propagation in Stations. PhD Thesis. TRAIL Thesis Series no. T2006/6, The Netherlands, 142 p.

Yuan, J., 2008. Railway timetable & traffic. Analysis, modelling, simulation. Chapter 10. Statistical analysis of train delays, 170-181.