Original article
DOI: 10.21683/1729-2646-2018-18-1-53-60

Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors[1]

Aleksey M. Zamyshliaev, *JSC NIIAS, Moscow, Russia*
Aleksey N. Ignatov, *Moscow Aviation Institute, Moscow, Russia*
Andrey I. Kibzun, *Moscow Aviation Institute, Moscow, Russia*
Evgeni O. Novozhilov, *JSC NIIAS, Moscow, Russia*


Aleksey M. Zamyshliaev


Aleksey N. Ignatov


Andrey I. Kibzun


Evgeni O. Novozhilov

Aim. *Rolling stock derailment is one of the most hazardous transportation incidents. Depending on the gravity of the consequences they may also be classified as accidents or train wrecks. The consequences of a derailment may vary from routine maintenance of the track and one or two wagons to an overhaul of the track and depot repairs of three or more wagons, as well as loss of cargo and long interruption of service. It must be noted that beside the damage to infrastructure and rolling stock caused by derailments there is a risk of environmental disaster. The Russian Federation along with the US, China and India has some of the world's longest rail networks that in places border with environmentally sensitive areas, e.g. national reserves and parks. Therefore, if a train carries hazardous cargo, e.g. gasoline or toxic gases and some of its wagons derailed, the harm related to the repair or decommissioning of rolling stock, track and possible loss of cargo may be aggravated by the damage caused by an environmental disaster that would cause great material losses to JSC RZD. In this context it appears to be of relevance to evaluate the functional dependency between the potential number of cars derailed and various factors, e.g. speed or amount of cargo carried by the train, for subsequent preparation of recommendations for the reduction of the potential number of derailed cars and, subsequently, reduction of possible harm.* **Methods.** *Probability theory and mathematical statistics methods were used, i.e. maximum likelihood method, negative binomial regression.* **Results.***For various groups of incidents, i.e. derailment as the result of wagon or locomotive unit malfunction out of switch, derailment as the result of rail malfunction out of switch, derailment at a switch not caused by previous derailment, specific functions of the average number of derailed wagons are identified. The paper shows a formula that allows – under a defined set of various factors, e.g. train speed, plan and profile of track, length and mass of the train – identifying the distribution series of the number of derailed wagons.* **Conclusions.** *The preliminary analysis of available Russian freight train derailment records it was shown that for various groups of transportation incidents the descriptive statistics of respective samples significantly differ, which is also the case for the US records. The construction of a functional dependence between the average number of derailed wagons and various traffic factors due to malfunction of wagons or locomotive units out of switches, it was identified that the available records do not suffice to forecast the number of derailed wagons in tangents. Mathematical models with a low superdispersion parameter were constructed for derailments due to track malfunction out of switches and derailments at switches.*

Keywords: *derailment, train wreck, traffic factors, maximum likelihood method, negative binomial regression.*

For citation: *Zamyshliaev AM, Ignatov AN, Kibzun AI, Novozhilov EO.Functional dependency between the number of wagons derailed due to wagon or track defects and the traffic factors. Dependability 2018;18(1): 53-60. DOI: 10.21683/1729-2646-2018-18-1-53-60*

[1] The deliverables were obtained as part of the performance of public assignment of the Ministry of Education of Russia no. 2.2461.2017/PCh and with the support of the Russian Foundation for Basic Research and JSC RZD within the framework of research project no. 17-20-03050 ofi_m_RZD

1. Introduction

The analysis of the factors that cause the most severe consequences of wagon derailments in train operations is based on the research of transportation incident records. For the period between 2013 and 2016 there are 262 records of derailment of freight wagons and passenger cars that occurred in the Russian Federation, not counting the records of transportation incidents classified as train wrecks. In the context of the railways of the US and India mentioned above out of [1, 2] follows that 2493 derailments occurred in the US in 2015 and 2016, while in India during the 2010-2011 and 2014-2015 periods 293 derailment took place, data for China is classified. It appears to be logical to use US records for the purpose of analysis, as it was done, for example, in the evaluation of damage caused by transportation incidents in [3], or Indian records. However, while the records of transportation incidents in the US and Russia are practically identical, differences exist. Russian records contain information on the presence or absence of switch at the location of derailment, plan and profile of track. We believe those factors bear upon not only the derailment itself, but the gravity of the consequences as well. There is no publicly available detailed information on the transportation incidents in India. Therefore American or Indian records cannot be used and the analysis will be based on Russian data only.

While researching the numbers of derailed wagons a confidence interval can be constructed for the potential number of derailed wagons of the distribution law of the number of derailed wagons can be deduced. However, those characteristics will be insufficient, as they will be identical for trains of both 3 wagons and 63 wagons. Therefore, the functional dependence between the number of derailed wagons and various factors must be identified. In [4], among the factors that have an effect on the consequences of derailments, the following ones are set forth: speed at the moment of derailment, remaining length (total number of wagons starting from the first derailed one), presence of additional locomotives in the middle/tail, proportion of loaded wagons. However, information of the curve radius and presence of gradient at the location of derailment was not taken into consideration. In [5] the level of derailment (level of hazard) was evaluated that depends on the class of track, tonnage handled, presence of signalling systems (e.g. train detection). The resulting dependence is integral in its nature and does not enable the reduction of the risk of derailment for individual trains. In [6] a similar task was researched that was related to finding the functional dependence between the probability of derailment and length of the train, number of kilometers travelled and class of track. However, the track geometry was not taken into consideration either. In this context the dependences proposed in [4-6] must be specified and clarified subject to the task under consideration in order to enable the development of practical recommendations.

The examination of Russian wagon derailment records has shown that some of them are not completely filled, i.e. there is the problem of missed data. Some values are missed both for the speed at the moment of derailment and the number of wagons in the trainset. As it is very difficult to recover those parameters, these observations were excluded from further analysis. Besides that, some derailments occurred not due to technical causes (track condition, bogie condition), but rather weather conditions or human factor that cannot be expressed in the nominal scale, hence such observations were not considered either. Consequently, various functional dependences in this paper were constructed based on 172 observations. Samples with and without missed data were compared as well.

2. Preliminary data analysis

First, let us construct a frequency diagram and find the descriptive statistics of the number of derailed cars.

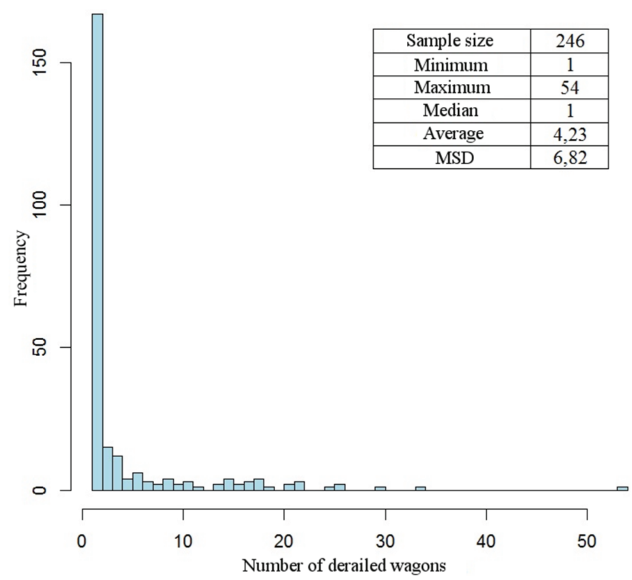

Figure 1. Frequency diagram and descriptive statistics of the number of derailed cars in case of freight train derailments and crashes

As we can see in figure 1, in most cases one wagon derails, while the average number of derailed wagons is about 4, while MSD is about 1.5 of the sample average. Therefore, it is important to find the functional dependence between the number of derailed wagons and the values of associated factors in order to reduce the severity of the consequences of derailments.

The descriptive statistics of the number of derailed wagons differ depending on the cause of derailment and the presence or absence of switch at the location of derailment.

So further analysis will be made for the three groups of accidents individually: derailment as the result of wagon or locomotive unit malfunction out of switch, derailment as the result of rail malfunction out of switch, derailment on a switch not caused by previous derailment due to track

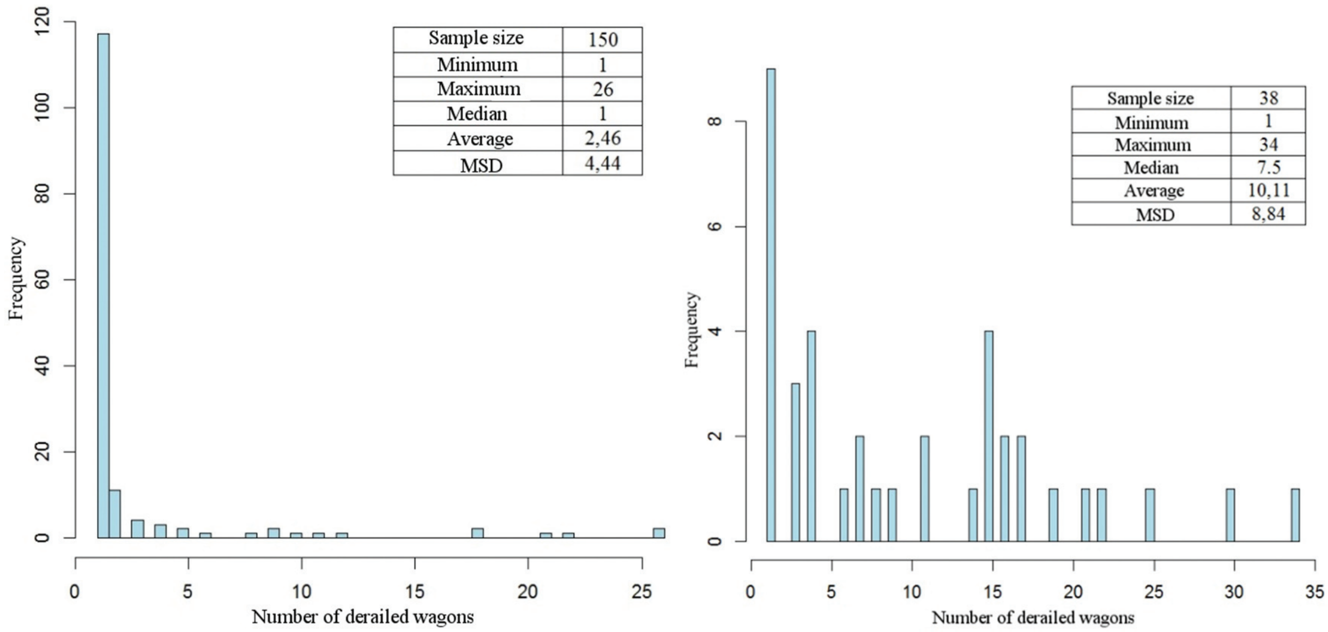

Figure 2. Frequency diagram and descriptive statistics of the number of derailed wagons in case of freight train derailments out of switches caused by wagon or locomotive unit (*left*) malfunction and track malfunction (*right*)

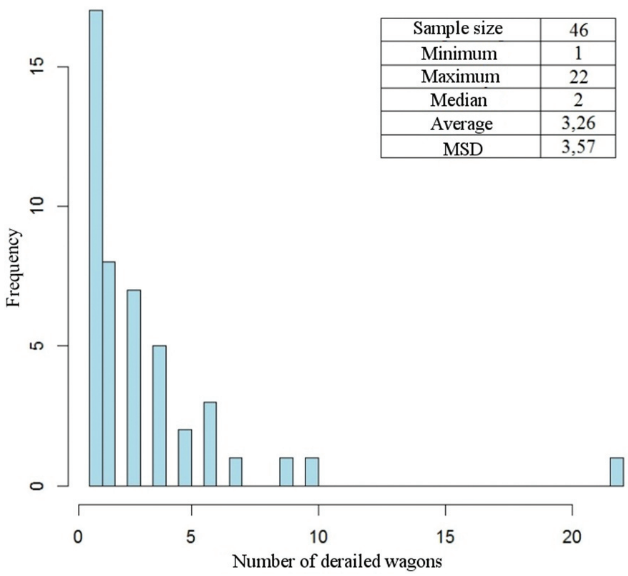

Figure 3. Frequency diagram and descriptive statistics of the number of derailed cars in case of freight train derailments at switches not caused by previous derailments, due to track or wagons/locomotive units malfunction

or wagon/locomotive unit malfunction. Note that within the time period under consideration 4 derailments occurred on switches as the result of a derailed wagon caused other wagons to derail after hitting the switch. These cases are classified separately, because derailed units do not always cause further derailments due to contact with s switch.

3. Primary designations

In the *j*-th group of incidents out of n_j transportation incident records involving freight train wagons derailment during train operation let us examine a certain *i*-th record. Let

c_{ij} be the total number of derailed units of rolling stock (locomotive sections and wagons);

k_{ij} bethe counting number of the unit (from the head of the train) that was the first to derail;

v_{ij} be the speed of the train at the moment of derailment, km/h;

l_{ij} be the number of wagons in the train;

l_{ij}^L be the total number of locomotive units in the train;

w_{ij} be the weight of the train, t;

$æ_{ij}$ be the rate of curve (value inversely proportional to the curve radius) at the place of derailment (for tangents the rate of curve is taken to be equal to zero);

γ_{ij} be the track profile at the place of derailment measured in promille having the minus sign if the gradient is downward and plus sign if the gradient is upward.

Let us also introduce an auxiliary variable $c^{\max}=l^L+l-k+1$, that is the realization of certain random value $C^{\max}=l^L+l-K+1$, where K is the random value that characterizes the number of the first derailed unit. Further, we will call random value $C^{\max}$ the remaining length of the train. Note that there is a statistical relation between the number of the derailed wagons and the remaining length of the train [4, 7-8]. Let us introduce another auxiliary variable (function) $\tilde{\mu}(w,l)$ that characterizes the loading factor (per [4]) of the train that depends on the train weight w and the number l of the transported wagons that is calculated using formula

$$\tilde{\mu}(w,l) \overset{def}{=} \pi_1 \frac{w}{l} + \pi_2, \qquad (1)$$

where π_1 and π_2 are unknown parameters. Consequently, the closer the function $\tilde{\mu}(w,l)$ is to zero, the higher is the number of empty wagons in the trainset. And vice versa, the closer the function $\tilde{\mu}(w,l)$ is to one, the lower is the number of empty and higher is the number of loaded wagons in the trainset. As the tare of a four-axle wagon is about 23 tons and the carrying capacity is around 69 tons, the coefficients π_1 and π_2 can be found by solving a system of linear equations

$$\begin{cases} 23\pi_1+\pi_2=0, \\ (23+69)\pi_1+\pi_2=1. \end{cases} \quad (2)$$

By solving system (2) we obtain π_1=1/69, π_2=–1/3. By substituting the obtained π_1 and π_2 into (1) we obtain

$$\tilde{\mu}(w,l)=\frac{w}{69l}-\frac{1}{3},$$

By setting $\tilde{\mu}_{ij}\overset{\text{def}}{=}\tilde{\mu}(w_{ij},l_{ij})$ we will obtain the train's weight ratio at the i-th derailment in the j-th group of incidents. Note that in the US incident records the number of loaded wagons is given explicitly, while in the Russian records there is not such characteristic, hence the ratio of loaded-to-total number of wagons has to be estimated.

In some records one or another characteristic may be missing or given inexplicitly which causes the problem of missed data. The missed values are often averaged out of the available ones, but in the context of the task at hand this approach cannot be used, as each transportation incident is unique and their number is not large. For that reason for each group of incidents we will further compare samples with missed values and complete sets of required characteristics.

4. Problem definition and method of solution

Let us examine the j-th group of transportation incidents, the total number of which within the period under consideration is n_j. Let C_j be a random value that characterizes the number of wagons and locomotive units that will derail as part of a group of incidents. As a derailment will inevitably involve not less than one unit of rolling stock, the following equality has place

$$C_j=1+\tilde{C}_j,$$

where $\tilde{C}_j$ is an auxiliary non-negative random value, of which the distribution law we will estimate later that has values in the set Z_+. Note that the distribution series of random values C_j and $\tilde{C}_j$ depend on the set of parameters $w,l,\tilde{\mu}(w,l),æ,\gamma$ and the realization $c^{\max}$ of random value $C^{\max}$, while the realizations $\tilde{c}_{ij}$ of random value $\tilde{C}_j$ can be obtained from formula $\tilde{c}_{ij}=c_{ij}-1$.

As random value $\tilde{C}_j$ is discrete, we cannot use the linear regression tools in the evaluation of the functional dependence between this random value and parameters $c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma$. Ordinal regression is partially similar to linear regression for an integral-valued dependent variable, yet in our case it cannot be used either, as not for all numbers out of the range of sample realization values there are derailments with identical numbers of derailed wagons.

Quantile regression [9] is another method of finding the desired dependence. However, due to the small number of observations at the level of dependability of, for example, α=0,999 and with 40 observations quantile regression does not appear to be usable. For that reason we will use the maximum likelihood method and negative binomial regression, namely we will assume conditional distribution $\tilde{C}_j$ to be common-negative binomial distribution with parameters r_j and p_j.

Let us recall the formulas for negative binomial regression for the case under consideration [10-11]

$$\begin{cases} P\{\tilde{C}_j=\tilde{c}\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma\}= \\ \quad =\dfrac{\Gamma(\tilde{c}+r_j)}{\Gamma(\tilde{c}+1)\Gamma(r)}p_j^{\tilde{c}}(1-p_j)^{r_j}, \\ \mathbf{M}[\tilde{C}_j\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma]=\dfrac{p_jr_j}{1-p_j}, \\ \mathbf{D}[\tilde{C}_j\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma]=\dfrac{p_jr_j}{(1-p_j)^2}. \end{cases} \quad (3)$$

Let

$$\mathbf{M}[\tilde{C}_j\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma]= \\ =f_j(a_{1j},a_{2j},\ldots,a_{m_jj},c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma), \quad (4)$$

$$\mathbf{D}[\tilde{C}_j\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma]= \\ =f_j(a_{1j},a_{2j},\ldots,a_{m_jj},c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)(1+ \\ +\theta_jf_j(a_{1j},a_{2j},\ldots,a_{m_jj},c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)). \quad (5)$$

where $f_j(a_{1j},a_{2j},\ldots,a_{m_jj},c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)$ is a function that is normally the exponential transformation of the linear function based on the parameters of function $a_{1j},a_{2j},\ldots,a_{m_jj}$ to be defined; parameter $\theta_j>0$ characterizes the superdispersion and is also to be defined.

By substituting (4) – (5) into (3) and introducing for convenience the additional designation $a_j\overset{\text{def}}{=}(a_{1j},a_{2j},\ldots,a_{m_jj})^{\mathrm{T}}$, we obtain

$$p_j=\frac{\theta_jf_j(a_j,c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)}{1+\theta_jf_j(a_j,c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)},\ r_j=\frac{1}{\theta_j}.$$

Thus

$$P\{\tilde{C}_j=\tilde{c}_{ij}\,|\,C^{\max}=c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma\}= \\ =\frac{\Gamma(\tilde{c}_{ij}+1/\theta_j)}{\Gamma(\tilde{c}_{ij}+1)\Gamma(1/\theta_j)}(\theta_jf_j(a_j,c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma))^{\tilde{c}_{ij}}\times \\ \times(1+\theta_jf_j(a_j,c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma))^{-(\tilde{c}_{ij}+1/\theta_j)}.$$

Let us construct the log-likelihood function

$$\overline{L}_j(a_j,\theta_j,\tilde{c}_{ij},c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij},i=\overline{1,n_j})=$$
$$=\ln\left(\prod_{i=1}^{n_j}P\{\tilde{C}_j=\tilde{c}_{ij}\mid c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij}\}\right)=$$
$$=-n_j\ln(\Gamma(1/\theta_j))+\sum_{i=1}^{n_j}(\ln(\Gamma(\tilde{c}_{ij}+1/\theta_j))+$$
$$+\tilde{c}_{ij}\ln(\theta_j f_j(a_j,c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij}))-$$
$$-(\tilde{c}_{ij}+1/\theta_j)\ln(1+\theta_j f_j(a_j,c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij}))-$$
$$-\ln(\Gamma(\tilde{c}_{ij}+1))).$$

Let us set the problem of finding the unknown vector a_j and parameter θ_j

$$(\theta_j^*,a_j^*)^{\mathrm{T}}=$$
$$=\arg\max_{a_j\in R^{m_j},\theta>0}\overline{L}_j(a_j,\theta_j,\tilde{c}_{ij},c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij},i=\overline{1,n_j}).(6)$$

Note that the quality of the constructed model (selection of function $f_j(\cdot)$) is characterized by not only the optimal value of the log-likelihood function $\overline{L}_j(a_j^*,\theta_j^*,\tilde{c}_{ij},c_{ij}^{\max},w_{ij},l_{ij},\tilde{\mu}_{ij},æ_{ij},\gamma_{ij},i=\overline{1,n_j})$, but the value of parameter θ_j^* as well. The closerparameter θ_j^* is to zero, the better is the constructed model, as the dispersion of random value $\tilde{C}_j$ is linear in parameter θ.

5. Solution of the problem

First, let us describe the general principles of selection of functions $f_j(\cdot)$. According to [9] functions should be selected with exponentials of a certain function linear in the evaluated regression parameter. According to [4] the selected factors that accompany a transportation incident are the logarithms of movement speed and remaining train length, loading factor, as well as their various combinations. Now let us consider each group of transportation incident individually.

5.1. Derailment due to wagon or locomotive unit malfunction out of switch

For this group of traffic incidents let us select function $f_1(\cdot)$ as follows

$$f_1(a_1,c^{\max},w,l,\tilde{\mu}(w,l),æ,\gamma)=f_1(a_1,c^{\max},\tilde{\mu}(w,l),æ,\gamma)=$$
$$=\chi_{æ\neq0}\exp\{a_{11}+a_{21}æ\ln(v)+a_{31}(1-\tilde{\mu})\ln(c^{\max})\ln(v)\cdot$$
$$\cdot\min(0,\gamma)+a_{41}(1-\tilde{\mu})\ln(c^{\max})+a_{51}(1-\tilde{\mu})^2\ln(c^{\max})\ln(v)\cdot$$
$$\cdot\min(0,\gamma)+a_{61}\tilde{\mu}+a_{71}\chi_{\gamma>0}\tilde{\mu}\ln(c^{\max})\ln(v)\}+$$
$$+\chi_{æ=0}\exp\{a_{81}+a_{91}\chi_{\gamma<0}\ln(c^{\max})\ln(v)+a_{101}\chi_{\gamma>0}\tilde{\mu}\ln(v)\ln^2(c^{\max})\},$$

where χ_A is the characteristic (indicator) function of a certain event A i.e.

$$\chi_A\overset{\text{def}}{=}\begin{cases}0, & x\notin A,\\ 1, & x\in A.\end{cases}$$

Let us comment the choice of function $f_1(\cdot)$. The function splits into two summands: the first one characterizes the gravity of consequences of derailment in curves ($\chi_{æ\neq0}$), while the second characterizes the gravity of consequences of derailment in tangents ($\chi_{æ=0}$).

In the part related to the derailments in curves three groups of summands can be identified: the first group contains the summands with parameters a_{11} and a_{21} that are invariant by the train load, the second group contains the summands with parameters a_{31}, a_{41}, a_{51} of which the effect increases with the reduction of train load, the third group contains the summands with parameters a_{61}, a_{71} of which the effect increases with the growth of train load. The severity of the consequences for loaded trains is increased by the presence of upward gradient, while for empty trains it is increased by not only the presence, but also the degree of downward gradient. The common trait of all the groups of summands is the fact that almost every summand there is either a logarithm of movement speed, or a logarithm of maximum number of derailed wagons, or sometimes their product. That is due to the fact that as the speed and maximum number of derailed wagons grows, a higher number of wagons are supposed to derail.

In the part related to derailments in tangents the summand with parameter a_{91} is not zero in case of movement along downward gradients, the summand with parameter a_{101} is not zero in case of movement along upward gradients. Derailments with serious consequences (more than 15 derailed wagons) happened not in steep downward or upward gradients, hence only the presence of a gradient rather than its degree is used.

By solving problem (6) we obtain the following estimates of maximum likelihood a_{11}^*, a_{21}^*, …, a_{101}^*, θ_1^*.

Note that in case of upward gradient in curves all the summands except the constant in function $f_1(\cdot)$ are non-negative under the obtained values of parameters a_{21}^*, …, a_{101}^*. Therefore, any increase of parameters $c^{\max}, v, \tilde{\mu}(w,l), æ$ causes a higher average number of derailed wagons, which corresponds with the physics of derailment. In case of downward gradient in curves there is a non-positive summand with parameter a_{61}^*. That is, among other things, due to the fact that in case of low train load $\tilde{\mu}<0{,}13587$ the sample average number of derailed trains was 4.54 wagons, while the sample average was 1.38 wagons in case of $\tilde{\mu}\geq0{,}13587$. This assumption is confirmed by the suggested model as

Table 1. Estimation of maximum likelihood a_{11}^*, a_{21}^*, …, a_{101}^*, θ_1^* based on sample with derailments in curves and tangents

a_{11}^*	a_{21}^*	a_{31}^*	a_{41}^*	a_{51}^*	a_{61}^*	a_{71}^*	a_{81}^*	a_{91}^*	a_{101}^*	θ_1^*
-7.76	315.69	286.88	0.63	-333.03	4.32	0.17	-1.55	0.2	0.04	3.83

Table 2. Estimation of maximum likelihood a_{11}^*, a_{21}^*, …, a_{101}^*, θ_1^* based on sample with derailments only in curves (65 observations)

a_{11}^*	a_{21}^*	a_{31}^*	a_{41}^*	a_{51}^*	a_{61}^*	a_{71}^*	a_{81}^*	a_{91}^*	a_{101}^*	θ_1^*
-7.25	307.22	284.43	0.54	-329.86	3.85	0.16	–	–	–	1.87

Table 3. Estimation of maximum likelihood a_{11}^*, a_{21}^*, …, a_{101}^*, θ_1^* based on sample with derailments only in angents (35 observations)

a_{11}^*	a_{21}^*	a_{31}^*	a_{41}^*	a_{51}^*	a_{61}^*	a_{71}^*	a_{81}^*	a_{91}^*	a_{101}^*	θ_1^*
–	–	–	–	–	–	–	-1.52	0.2	0.04	6.05

$$\frac{\partial f_1(a_1^*, c^{\max}, \tilde{\mu}(w,l), \ae, \gamma)}{\partial \gamma} =$$
$$= f_1(a_1, c^{\max}, \tilde{\mu}(w,l), \ae, \gamma)(a_{31}^*(1-\tilde{\mu})\ln(v)\ln(c^{\max}) +$$
$$+ a_{51}^*(1-\tilde{\mu})^2 \ln(v)\ln(c^{\max})) < 0 \Leftrightarrow a_{31}^* + a_{51}^*(1-\tilde{\mu}) < 0 \Leftrightarrow$$
$$\Leftrightarrow 1-\tilde{\mu} < -\frac{a_{31}^*}{a_{51}} \Leftrightarrow \tilde{\mu} > 1 + \frac{a_{31}^*}{a_{51}} = 0,13587.$$

Note that records regarding derailments in tangents are insufficient, as if we examine the samples on the derailments in tangents and curves separately, the results will be as follows.

By comparing the results in tables 1, 2 and 3 we conclude that joint consideration of derailments in curves and tangents somewhat alters the predicted average number of derailed wagons in curves, while leaving the average number of derailed wagons in tangents practically unchanged. At the same time the dispersion that is characterized by parameter θ_1^* changes significantly. Therefore, derailments in curves and tangents should be separated from each other.

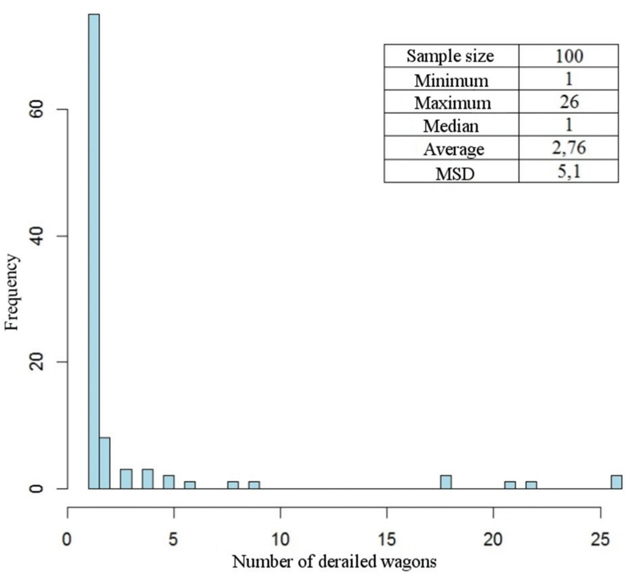

Figure 4. Frequency diagram and descriptive statistics of the number of derailed wagons in case of freight train derailments out of switches caused by wagon or locomotive unit malfunction according to a sample with no missed data

Additionally, during data processing it turned out that the sample average of the number of derailed wagons in tangents is higher than the sample average of the derailed wagons in curves (3.74 versus 2.23 wagons). Therefore, additional research is required both in terms of the depth of research (increased number of considered records) and in terms of the quality of considered records, namely the clarification of information on the causes of the occurred accidents and track characteristics in the location of derailment, especially in tangents.

A detailed example of the resultant formulas is given in [12].

In this section, the analysis was based on a sample with the following characteristics.

5.2. Derailment due to track malfunction out of switch

For this group of traffic incidents let us select function $f_2(\cdot)$ as follows

$$f_2(a_2, c^{\max}, w, l, \tilde{\mu}(w,l), \ae, \gamma) = f_2(a_2, c^{\max}, \tilde{\mu}(w,l), \ae, \gamma) =$$
$$= \exp\{a_{12} + a_{22}\tilde{\mu} + a_{32}\ln(v) + a_{42}\ln(c^{\max})\}.$$

The principle of function $f_2(\cdot)$ construction is similar to the one of function $f_1(\cdot)$. This function is also similar to the one suggested for the estimation of the average number of derailed wagons due to track malfunction in [4]. Additionally, in function $f_2(\cdot)$ unlike in $f_1(\cdot)$ the is no parameter γ. That is due to the fact that out of 38 incidents caused by track malfunction in 11 cases it was impossible to identify the gradient value. Parameter æ is also absent as it was used in the identification of the model with the best log-likelihood function value.

By solving problem (6) we obtain the following estimates of maximum likelihood a_{12}, a_{22}, a_{32}, a_{42}, θ_2^*.

Table 4. Estimated maximum likelihood a_{12}^*, a_{22}^*, a_{32}^*, a_{42}^*, θ_2^*

a_{12}^*	a_{22}^*	a_{32}^*	a_{42}^*	θ_2^*
-6.4	1.01	0.68	1.48	0.3

All the summands of function $f_2(\cdot)$ turn out to be positive and therefore any increase in the traffic parameters will cause a higher average number of derailed wagons, which is logical.

In this section, the analysis was based on a sample with the following characteristics.

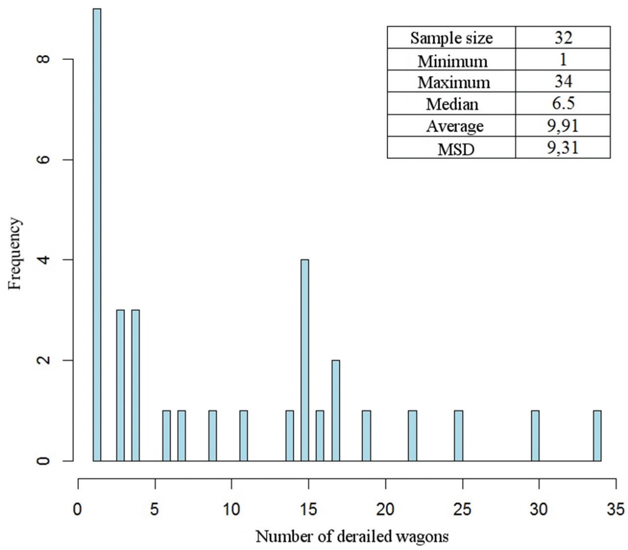

Figure 5. Frequency diagram and descriptive statistics of the number of derailed wagons in case of freight train derailments out of switches caused by track malfunction according to a sample with no missed data

5.3. Derailment at a switch not caused by prior derailment

For this group of traffic incidents let us select function $f_3(\cdot)$ as follows

$$f_3(a_3, c^{\max}, w, l, \tilde{\mu}(w,l), æ, \gamma) = f_3(a_3, c^{\max}, \tilde{\mu}(w,l), æ, \gamma) =$$
$$\exp\{a_{13} + a_{23}\tilde{\mu}\ln(v) + a_{33}\tilde{\mu}\ln^2(c^{\max}) +$$
$$+a_{43}\ln(v) + a_{53}\ln(v)\ln(c^{\max})\}.$$

Note that in this case function $f_3(\cdot)$ does not contain variables γ and æ, as it is extremely difficult or sometimes even impossible to identify them for incidents that occurred at switches.

By solving problem (6) we obtain the following estimates of maximum likelihood a_{13}, a_{23}, ..., a_{53}, θ^*_3.

Table 5. Estimated maximum likelihood a^*_{13}, a^*_{23}, ..., a^*_{53}, θ^*_3

a^*_{13}	a^*_{23}	a^*_{33}	a^*_{43}	a^*_{53}	θ^*_3
-1.49	0.99	-0.16	-0.91	0.43	0.41

In this section the analysis was based on a sample with the following characteristics.

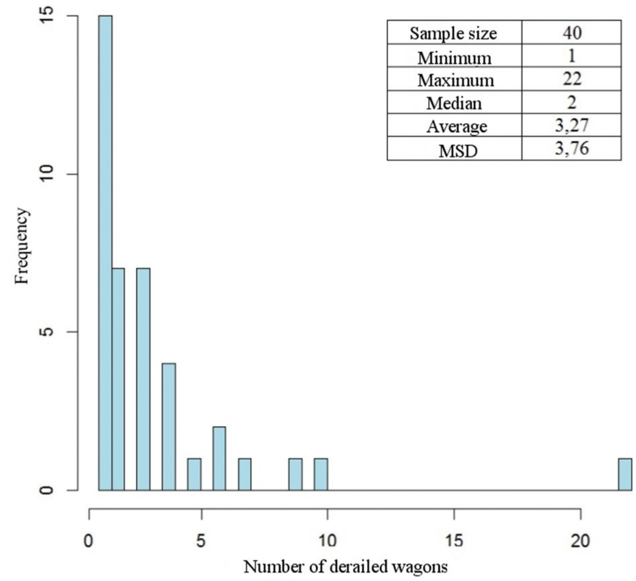

Figure 6. Frequency diagram and descriptive statistics of the number of derailed cars in case of freight train derailments at switches not caused by previous derailments

6. Conclusion

This paper shows a functional dependency between the average number of derailed wagons and various traffic factors: train speed, plan and profile of track, length and mass of the train. Various groups of transportation accidents are defined: derailment as the result of wagon or locomotive unit malfunction out of switch, derailment as the result of rail malfunction out of switch, derailment on a switch not caused by previous derailment. Based on the maximum likelihood method and negative binomial regression, functions of average number of derailed wagons are defined. The paper shows a formula that allows – under a defined set of various factors, e.g. train speed, plan and profile of track, length and mass of the train – identifying the distribution series of the number of derailed wagons. The results of the research can later be used in the evaluation of the risk of freight train derailment throughout the Russian rail network.

References

1. <http://safetydata.fra.dot.gov/>
2. <http://www.indianrailways.gov.in/>
3. Goriainov AV, Zamyshliaev AM, Platonov EN. Analysis of effects of factors on the damage caused by accidents in transportation using regression models. Dependability 2013;2:126-144.
4. Liu X, Saat MR, Qin X, Barkan CPL. Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. Accident Analysis and Prevention 2013;59:87-93.
5. Liu X, RapikSaat M, Barkan CPL. Freight-train derailment rates for railroad safety and risk analysis. Accident Analysis and Prevention 2017;98:1-9.

6. Anderson RT, Barkan CPL. Derailment probability analysis and modeling of mainline freight trains. In: Proceedings of the 8-th International Heavy Haul Conference, Rio de Janeiro (Brazil): International Heavy Haul Association; 2005. p. 491–497.

7. Bagheri M, Saccomanno F, Chenouri S, Fu LP. Reducing the threat ofin-transit derailments involving dangerous goods through effective placement along the train consist. Accident Analysis and Prevention 2011;43:613–620.

8. Saccomanno FF, El-Hage S. Minimizing derailments of railcars carrying dangerous commodities through effective marshaling strategies. Transportation Research Record 1989;1245:34–51.

9. Koenker R, Hallock KF. Quantile Regression. Journal of Economic Perspectives 2001; 15(4):143-156.

10. DeGroot MH, Schervish MJ. Probability and Statistics. 4th ed. Addison-Wesley; 2012.

11. Cameron AC, Trivedi PK. Essentials of Count Data Regression in A Companion to Theoretical Econometrics. Baltagi BH, editor. Blackwell Publishing Ltd; 2003.

12. Zamyshliaev AM, Ignatov AN et al. Ob otsenkekolichestvavagonov v skhodepripoezdnoyrabotenaosnovefaktornykhmodeley [On the evaluation of the number of derailed wagons in operation based on factor models]. In: Proceedings of the Sixth Science and Engineering Conference Intelligent Control Systems in Railway Transportation. Computer and mathematical modeling. ISUZHT 2017. 2017. p. 132-135.

About the authors

Aleksey M.Zamyshliaev, Doctor of Engineering, Deputy Director General, JSC NIIAS, Moscow, Russia, phone: +7 495 967 77 02, e-mail: A.Zamyshlaev@gismps.ru

Aleksey N. Ignatov, Moscow Aviation Institute, postgraduate student, Moscow, Russia, phone: +7 906 059 50 00, alexei.ignatov1@gmail.com

Andrey I. Kibzun, Doctor of Physics and Mathematics, Professor, Moscow Aviation Institute, Head of Chair, Moscow, Russia, phone: +7 499 158 45 60, e-mail: kibzun@mail.ru

Evgeny O. Novozhilov, Candidate of Engineering, Head of Unit, JSC NIIAS, Moscow, Russia, phone: +7 495 967 77 02, e-mail: eo.novozhilov@vniias.ru.

Received on 16.10.2017