Forecasting The Spread of the Coronavirus Infection to Consider Forces During Emergency

Elena Arefieva Valentinovna [1,2][a], Elena Muravyeva Viktorovna [2][b]
and Alsu Shakirova Ilnurovna[2][c]
[1]FSBI All-Russian Research Institute on Civil Defense and Emergency Situations of The Ministry of Emergency Situations, Davydkovskaya str., Moscow, Russia
[2]Department of Industrial and Environmental Safety, Kazan National Research Technical University named after A.N. Tupolev-KAI, K.Marx str., Kazan, Russia
elaref@mail.ru, elena-kzn@mail.ru, alsugal92@mail.ru

Keywords: Forecasting, The spread of the coronavirus infection, Emergency, Mathematical modelling, The pandemic.

Abstract: The pandemic has identified additional requirements for rescue units and other services of the emergency department in the field of ensuring the protection of the population from natural and man-made hazards. In this regard, it became necessary to assess the number of capable employees of the disaster response force, taking into account the number of sick employees. To establish a regressional dependence between random variables characterizing the number of able-bodied employees and the number of sick from the number of the disaster response forces, the authors propose to use transformed and appropriately normalized random variables characterizing capable and sick employees.

1 INTRODUCTION

In the modern world, almost everyone has already heard about the global epidemic of the new coronavirus COVID-19. There are thousands of people infected with this virus in almost every country. The number of cases is growing every day. Modern media publish daily statistics and information about the virus, but it is not always reliable.

Coronaviruses can cause a number of diseases in humans – from mild forms of acute respiratory infection to severe acute respiratory syndrome. Currently, it is known about the circulation of four types of coronaviruses among the population (HCoV-229E, -OC43, -NL63 and -HKU1), which are present all year round in the structure of acute respiratory viral infections, and, as a rule, cause damage to the upper respiratory tract of mild to moderate severity.

Currently, data on the epidemiological characteristics of a new coronavirus infection caused by SARS-CoV-2 are limited.

A person infected with the COVID-19 coronavirus does not suspect his illness for a long period, actively contacts others, and by the time signs of the disease appear, he manages to infect a significant number of people. WHO has calculated the incubation period (the period of time between infection and the appearance of clinical symptoms of COVID-19 virus disease) ranges from 1 to 14 days and is most often about five days. The infection is transmitted by airborne droplets (when coughing, sneezing, talking), dust and nonpercutaneous channels. Transmission factors are air, food and household items contaminated with SARS-CoV-2.

As shown in Figure 1 of the Coronavirus biology project, the base reproductive number of COVID-19 is much higher than that of Ebola, which had an outbreak in 2014-2015 and the Spanish influenza of the early 20th century.

[a] https://orcid.org/0000-0001-6037-9663
[b] https://orcid.org/0000-0002-0547-6417
[c] https://orcid.org/0000-0002-5977-8883

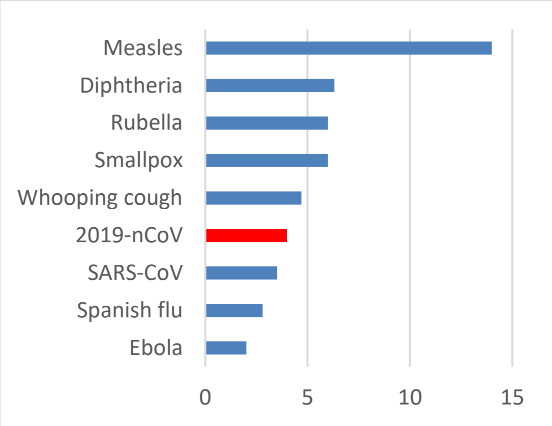

Figure 1: Basic reproductive number for various diseases.

During a pandemic, the involvement of forces to respond to disasters and conduct a possible evacuation of the population must be carried out in compliance with strict sanitary standards in the placement, food, water supply of the population (Sakhno, 2003; Wisner, 2002). The occurrence of an emergency situation caused by natural or man-made causes in the conditions of the spread of coronavirus infection requires the development of special models for calculating the necessary response forces, taking into account the disease, as well as employees of the disaster response forces.

To solve such problems, it is necessary to take into account:

- forecast of the situation that may develop as a result of a pandemic;
- forecast of the population that may suffer as a result of a natural disaster in the context of the spread of coronavirus infection;
- forecast of sanitary losses and temporary absence due to illness among employees of control bodies, personnel of emergency rescue teams.

To solve the above problems, it is necessary to use various mathematical models both to assess the spread of coronavirus infection and to assess changes in the composition of the response forces as a result of infection of employees.

The article proposes an approach to establish a regressional dependence between the number of disaster response service employees who are able to perform their work functions and employees who fall ill and cannot perform their work duties.

2 REFERENCES

Based on the Baroyan- Rvachev models in the 60-70s in the Soviet Union, the Institute of Experimental Medicine named after N.F. Gamaleya of the Academy of Medical Sciences of the USSR developed the unique models of the influenza epidemic for the entire territory of the USSR. By analogy with the movement of flows of energy, matter in the processes of hydrodynamics, the balance equations of the "flows" of individuals passing through the main stages of the infectious process were compiled: S - "susceptible"; E – "exposed", I – "infectious patients" - R – "recovered". Subsequently, the model was modified by scientists from different countries, depending on a particular epidemic, and has now received the world-famous abbreviation – SEIR (Baroyan, 1977).

Thus, the SEIR model operates with "flows" of individuals:

1. Susceptible S(t) – individuals without immunity.
2. Latent E(t) - individuals who have become spreaders of the disease after contact with infected people, but have not yet become ill (the incubation period of the disease).
3. Infected I(t) - diseased individuals spreading the disease.
4. Unsusceptible R(t) - individuals who have had a disease and received immunity.
5. Dead D(t) - individuals who died as a result of the disease.

It is worth noting that the spreaders of the disease are both those who are directly ill, and individuals who are carriers of the disease during the incubation period. The "flows" of individuals, i.e. the transitions of individuals from one state to another, can be schematically represented in the form (Figure 2):

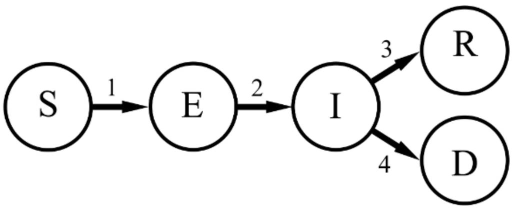

Figure 2: the spread of the disease, where 1 is an infection as a result of contact with an infected person, 2 is the onset of symptoms of the disease, 3 is recovery and acquisition of immunity, 4 is death due to illness; S(t) is susceptible, E(t) is latent; I(t) is sick (infected); R(t) is recovered and acquired immunity; D(t) is dead individuals.

The basic model for describing the spread of epidemics is the SIR model (S - susceptible, I - infected, R - recovered), proposed by Scottish epidemiologists Kermak and Mackendrick in the 1920s (Kermack, 1927). In the future, the model of the influenza epidemic proposed by Soviet scientists O.V. Baroyan and L.A. Rvachev (Baroyan, 1977). It became known as the Baroyan-Rvachev model. The method used in this model is based on the use of the mathematical apparatus of continuum mechanics (Landau, 1954), which simulates the process of transfer of matter (energy, momentum, etc.), and which can be applied to simulate the transfer of an infectious agent from sick to healthy. The influenza epidemic model is described by a system of nonlinear differential equations with corresponding initial and boundary conditions similar to the hydrodynamic equations (Landau, 1954).

Risk assessment approaches are widely used to make decisions on preventing and mitigating the consequences of the epidemic, including assessing the risk of spreading the pathogen, assessing the level of exposure and vulnerability, as well as assessing readiness to counteract a dangerous virus (Colizza, 2007). The quantitative assessment of the danger includes the determination of the quantitative characteristics of the pathogen (lethality, transmission routes, incubation period, etc.). The exposure assessment measures the dose of the pathogen that a person inhales, swallows, or comes into contact with. A database and information on the necessary resources in the field of biological risk assessment are publicly available on the special online portal of the Center for Advanced Microbiological Risk Assessment.

An interesting approach is based on a multiple-criteria decision analysis (MCDA), which uses a variety of criteria of an epidemiological, socio-economic nature (Vorobiev, 2015). This approach is also used by the European Center for Disease Prevention and Control (ECDC). This approach was applied in the development of WHO for the prevention and avoiding of epidemics, the definition of priority lists of the most dangerous diseases. WHO also uses a toolkit called the Global Burden of Disease (GBD) method, which is a scientific study to quantify the comparative magnitude of health loss caused by disease, injury, risk factors related to gender, geographic, socio-economic location. The work (World Health Organization, 2012) describes the approach and methodology for rapid risk assessment and provides operational guidelines for assessing the risk of disease spread. Attention is drawn to the fact that in order to perform a rapid risk assessment, including in the epidemiological field, it is necessary to have access to a database on diseases, a list of research results, and access to relevant literature.

The article (Chandra, 2020) presents a simulation of the spread of the epidemic caused by coronavirus infection in various environments using a stochastic model (susceptible-infected-recovered) (SIR), which also used Bayes' calculations.

At the same time, the use of an improved model using differential equations of the type (SIR) [14] requires a large amount of varied data (the number of individuals in the latency period, the number of sick, hospitalized, recovered, in critical condition, and deaths, as well as the time spent by people in these states, and other parameters) and does not allow an estimate of the number of able-bodied employees of the disaster response force depending on the sick employees. Therefore, this article attempts to describe an approach for estimating the number of response force personnel capable of performing their work functions depending on the sick employees, which is discussed in the works of S.N. Zaripova (Zaripova, 2009).

3 MODELS AND METHODS

Consider two random variables: *X* is the number of sick workers of Response Service, *Y* is the number of workers who continue to work in disaster response services, represented by the corresponding statistical series. After the initial statistical processing of the data, which consists in first constructing variational series consisting of ordered values of random variables *X* and *Y*, then statistical series consisting only of various sample values of random variables *X* and *Y*, and determining the frequencies of various sample values. The histograms show the frequencies of different values of the random variables *X* and *Y*. The constructed histograms (frequency polygons are constructed in a similar way) allow us to make an assumption about the types of distribution of the random variables *X* and *Y*.

If it is impossible to obtain an unambiguous answer about the types of distribution of random variables *X* and *Y* from the histograms and frequency polygons, then the compliance with the most common distribution laws (normal, logarithmically normal, Poisson's law, binomial) is checked using the χ^2 criterion (chi square). χ^2 – the criterion of agreement for testing the hypothesis about the distribution law of the random variable under study. If the observed values of χ^2_{obs} turned out to be greater than the critical values of χ^2_{cr} for all the distribution laws listed above, then transformations of the initial values of the random variables X and Y are performed. Before the transformation, the initial values are normalized according to the formula:

$$z_i' = (z_i - \overline{z})/\sigma_z \quad (1)$$

where $\overline{z}$ is the average value of the sample, σ_z is standard deviation.

Then the normalized values are converted according to the formula:

$$u_i' = lg(u_i + a) \quad (2)$$

where a - constant at which $u_i + a > 0$ for any i.

Further, for the transformed random variables, frequency histograms are constructed and correspondence to a certain (for example, normal) distribution law is established. Then the check is performed again by the χ^2 criterion (chi square).

4 RESULTS

Consider a random variable X is the number of sick workers, a random variable Y is the number of workers who continue to work in disaster response services in the period from 03/20/2021 to 05/31/2021 in the territory of the Republic of Tatarstan (Zaripova, 2007). From 73 elements of each of the samples X and Y, 20 and 12 different sample values were selected, respectively. Using the standard software package "Data Analysis", statistical series of samples and their graphical representations were obtained (Figure 3-4).

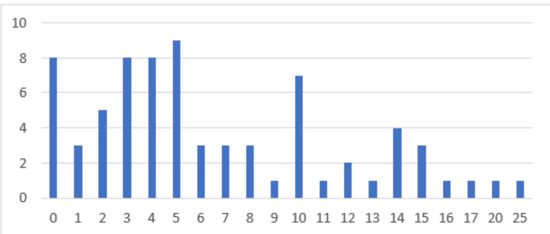

Figure 3: Histogram of random variable frequencies (X).

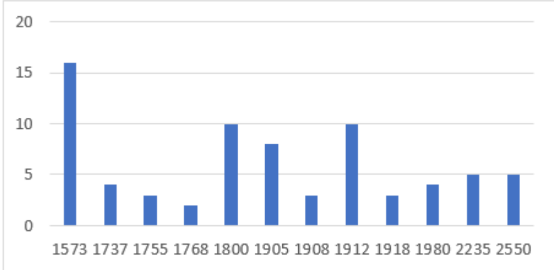

Figure 4: Histogram of random variable frequencies (Y).

Further, for each sample, the relative and accumulated relative frequencies of the sampled values are calculated. The polygons of the relative and accumulated frequencies are shown in Figure 5-8.

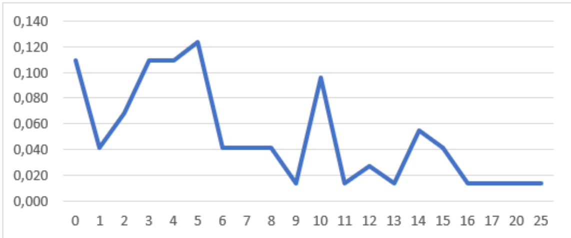

Figure 5: Polygon of relative frequencies of a random variable (X).

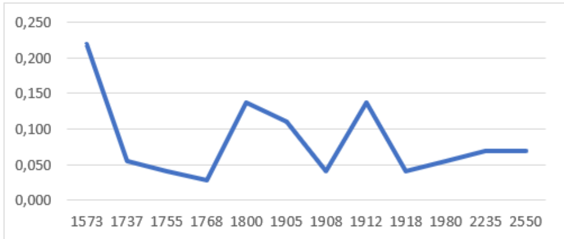

Figure 6: Polygon of relative frequencies of a random variable (Y).

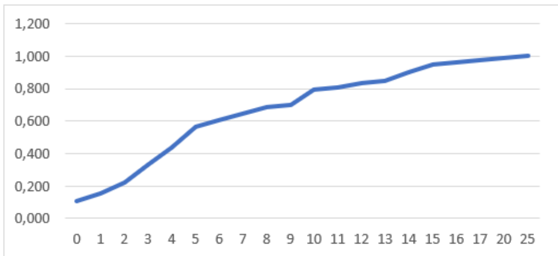

Figure 7: Polygon of accumulated frequencies of a random variable (X).

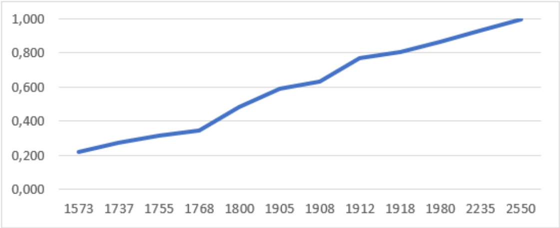

Figure 8: Polygon of accumulated frequencies of a random variable (Y).

Hypotheses about the distribution of random variables X and Y according to the normal law, logarithmically normal law, Poisson's law, binomial law were tested using statistical functions of the *MS Excel* library of built-in functions.

It is established that the observed values of random variables are much larger than the critical values of χ^2_{cr}, therefore, hypotheses about the distribution of random variables *X* and *Y* according to the normal law, logarithmically normal law, Poisson's law, binomial law contradict statistical data (Table 1).

$$\chi^2_{obs} = \sum_{i=1}^{k} \frac{(n'_i - n_i)^2}{n'_i} \tag{3}$$

where n_i is sample value frequency x_i or y_i, and $n'_i = N \cdot p_i$, where p_i - theoretical probability of sampled value x_i or y_i, N – sample size.

Table 1: χ^2_{obs} and χ^2_{cr} values for different distribution laws

The distribution law of a random variable	Normal		Logarithmically normal		Poisson		Binomial	
	X	Y	X	Y	X	Y	X	Y
χ^2_{obs}	138,24	164,5	47,52	26,15	98,76	27,45	143,66	17,19
χ^2_{cr}	28,87	11,07	28,87	11,07	28,87	11,07	28,87	11,07

Further, in accordance with formulas (1) and (2), transformations of the random variables *X* and *Y* are performed. As a result of such transformations, the formulas (4) and (5) were obtained for the random variable *X:*

$$x'_i = (x_i - 6{,}62)/5{,}38 \tag{4}$$

(rationing: $x_i \to x'_i$) and $x''_i = lg(x'_i + 4.3)$ $(x'_i \to x''_i)$, (5)

Substituting the value x'_i into formula (4) we obtain (6):

$$x''_i = lg(x_i + 16{,}51)/5{,}38 \tag{6}$$

Similar transformations are performed with the values of the random variable *Y*:

$$y'_i = (y_i - 1871{,}23)/254{,}32 \tag{7}$$

$$y''_i = lg(y'_i + 3{,}18) \tag{8}$$

$$y''_i = \frac{lg(y_i - 1062{,}78)}{254{,}32} \tag{9}$$

As a result, new values x''_i and y''_i of random variables were obtained, for which the histograms of frequencies were again constructed, according to which we can assume the presence of a normal distribution law (Figure 9-10).

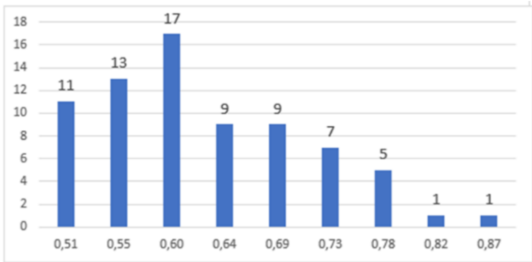

Figure 9: Histogram of converted frequencies X.

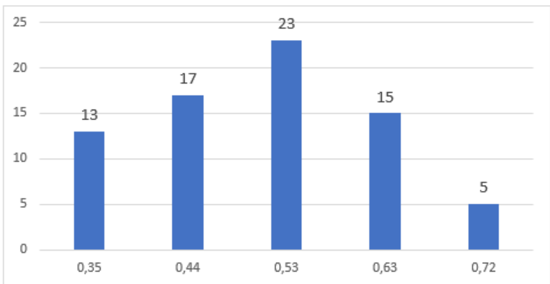

Figure 10: Histogram of converted frequencies Y.

The verification of this assumption is confirmed by the test according to the criterion χ^2 (chi square): at the significance level α = 0.05, the critical values of χ^2 - statistics, respectively, are equal to $\chi^2_{cr.X}$ = 12,59, $\chi^2_{cr.Y}$ = 7,81.

These values are greater than the corresponding calculated values $\chi^2_{calc.X}$ = 6,51 and χ^2_{calcY} = 6,63. Consequently, the null hypothesis H_0, which consists in the fact that the number of employees of disaster response services on the territory of the Republic of Tatarstan performing labor functions and persons from among the disaster response forces recognized as temporarily disabled (on sick leave) in the period from 03/20/2020 to 05/31/2020, obey the normal distribution law, does not contradict statistical data.

Based on the available values of the transformed random variables, we construct a linear regression equation, which has the form (10):

$$y'' = -0.3 + 1{,}259\, x'' \tag{10}$$

Regression statistics are summarized in Table 2, the graph of linear regression is shown in Figure 11.

Table 2: Regression statistics

Standard deviation b"=0,06	Standard deviation a"=0,03
Determination coefficient R^2 = 0.88	Standard deviation y=0,04
F- statistics = 518,77	Number of degrees of freedom = 71
Regression sum of squares = 1,03	Residual sum of squares = 0,14

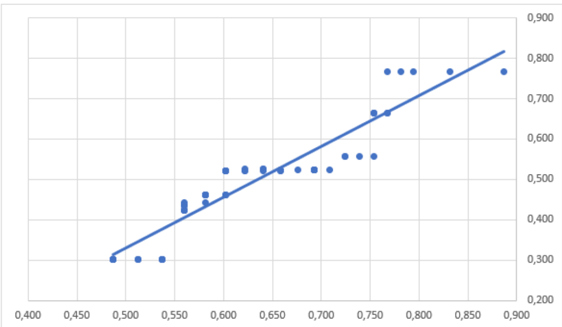

Figure 11: Graph of linear regression of transformed random variables.

The constant term and the regression coefficient are significant, since the corresponding probabilities of random events equal to 1,09E-12 and 2,3E-34 are less than the significance level α = 0,05. The average approximation error E = 7,71% is less than 8-10%, therefore model (10) is more accurate.

To pass to the initial values of random variables X and Y in equation (10) instead of x" and y" we substitute their values from (6) and (9), and after small transformations we obtain a dependence of the form (11):

$$y = 1032{,}78 + 254{,}32 \cdot 10^{-0{,}3} \times \times \left(\frac{x + 16{,}51}{5{,}38}\right)^{1{,}26} \tag{11}$$

on the basis of which it is possible to make forecasts of the number of employees of the disaster response services performing labor functions in the workplace, based on the values of the number of persons from among the employees of the response services recognized as temporarily disabled due to coronavirus infection.

To predict the values x = x(t), where t is time, on the basis of statistical data for 73 days, trend lines (polynomial second degree, exponential and linear) were constructed (Figure 12), equations, coefficients of determination and other characteristics of which are summarized in Table 3.

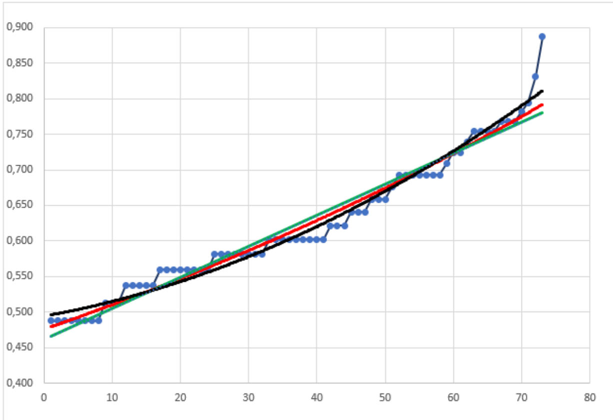

Figure 12: Trendlines.

Table 3: Trendline equations and their characteristics

Trends	Polynomial	Exponential	Linear
Trendline equation	$x = 0{,}00004t^2 + 0{,}0017t + 0{,}4942$	$x = 0{,}4758 \cdot e^{0{,}007t}$	$x = 0{,}0044t + 0{,}4609$
Determination coefficient, R^2	0,9837	0,9717	0,9508

5 DISCUSSIONS

According to the predicted values $x = x(t)$, obtained from one of the equations in Table 3, it is possible to predict the values of *y(x)* by formula (11), taking into account which it is possible to plan the actions of the disaster response forces.

To obtain a regression dependence of the type of dependence represented by formula (11), transformations of the original random variables were performed, then normalization was performed, which made it possible to obtain a regression dependence between the transformed values, and then using reverse substitution to obtain the required regression dependence.

Based on this approach, an assessment of the required number of employees in a pandemic can be carried out to prevent forecasted natural and man-made emergencies.

6 CONCLUSIONS

The obtained regression dependence of random variables (11) characterizing the number of able-bodied employees and sick employees, along with other computational models, predicting the spread of infection (increase in the number of infected) allow individuals to make more informed decisions on the configuration of emergency rescue forces, fire brigades in the event of emergencies, man-made accidents and fires (Arefeva, 2019).

REFERENCES

Baroyan, O.V., Rvachev, L.A., 1977. *The book*, Mathematics and epidemiology.

Sakhno, V.I., Zakharov, G.I., Karlin, N.E., Pilnik, N.M., 2003. *The book*, Organization of medical care for the population in emergency situations: a training manual. St. Petersburg, Publishing house FOLIO.

Wisner, B., Adams, J., 2002. *The book*, Environmental health in emergencies and disasters: a practical guide. World Health Organization, 1st edition.

Kermack, W.O., McKendrick, A.G. , 1927. A contributions to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* A,115. pp.700–721.

Baroyan, O.V., Rvachev, L.A., and Ivannikov, Yu.G., 1970. *The book*, Modelling and prediction of influenza epidemics in the USSR. Moscow, Institute of Epidemiology and Microbiology named after N.F. Gamalei.

Landau, L.D., Lifshits, E.M., 1954. *The book*, Continuum mechanics. Moscow, Gostekhizdat.

Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A., 2007. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study, BMC Med., 5:34.

Vorobiev, P.A., Kholovnya, M.A., Krasnova, L.S., 2015. Multiple-criteria decision analysis (MCDA) in Health Technology Assessment. *Pros and Cons. Pacific Medical Journal*, 1:1.pp.35-40.

World Health Organization, 2012. The book, Rapid risk assessment of acute public health events. World Health Organization. Geneva, WHO.

Vedant, C., 2020 "Stochastic Compartmental Modelling of SARS-CoV-2 with Approximate Bayesian Computation," *medRxiv*.

Zaripova, S.N., Kvaginidze, V.S., 2009. Noise and vibration safety of mining and transport equipment. *Mining information and analytical bulletin*, (10). pp. 309-317.

Zaripova, S.N., Kvaginidze, V.S., Petrov, V.F., 2007. Prediction of the safe functioning of excavator-automobile complexes of mining enterprises of the North. Moscow: *Publishing house of Moscow State Mining University.*

Arefeva, E.V., Muraveva, E.V., 2019. Forecast of the impact of hazardous factors of emergencies on the state of the objects of the technosphere. *Bulletin of the State Budgetary Institution. Scientific Center for Life Safety*, 2:40. pp. 62-67.

Arefeva, E.V., Muraveva, E.V. Frose, T.Yu., 2019. Considering emergency hazards in construction and operation of infrastructures," IOP Conf. Ser.: Mater. Sci. Eng. (687), 066023.