

# Visualization of the Dynamics of Changes in Structured Data by the Example of Covid-19 Development\*

Anna Trubakova<sup>[0000-0003-0280-1760]</sup> and Andrey Trubakov<sup>[0000-0003-0058-1215]</sup>

Bryansk State Technical University, Bryansk, Russia  
trubakovaaa@gmail.com, trubakovao@gmail.com

**Abstract.** It is very important to visualize the initial data in a decision-making system with human participation in a proper way. Decision-making complexity, the correctness of the decisions made often depend on it. Nowadays there are many different data visualizations today such as graphs, bar charts, pie and histograms, etc. However, the situation becomes more difficult, if it comes to visualizing the dynamics of changes of structured data. Structured data is understood in this paper as data in which the observed value has an internal structure and consists of a large number of components (characteristics). This paper provides a formal description of structured data and proposes an approach to displaying the graph of their change, by which one can judge not only the general process of the observed value, but also the change in the internal structure (components included in it and their influence on human decision-making). The area of potential application of this approach, as well as its structure features and capabilities are demonstrated in the work. The description of the developed software is given for modeling the situation with the spread of COVID-19 in Bryansk region where the proposed version of visualization is used. Examples of specific situations are considered in which this approach turned out to be useful and helpful. The possibility to use the proposed method in other applied areas is described at the end of the work.

**Keywords:** Data Visualization, Structured Data, Making Decisions, SIR Model.

## 1 Introduction

It is an obvious fact that the amount of data that needs to be analyzed when making management decisions greatly grows every year. At the same time, man remains responsible for the analysis and final decision, that is, they depend on the decision-maker (DM). Modern decision support systems (DSS), based on mathematical methods, can partially replace human resources at the stage of processing initial data, but not at the stage of final decision-making. The lack of clear and structured information during the

---

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

\* Publication is supported by RFBR grant № 19-07-00844.

decision-making process makes it difficult, due to processing data of various nature, frequency of information updates over a short period of time, while the support and decision-making is always left to DM. Due to this need for operational research, it is required to provide high-quality visualization of constantly growing and developing volume of data that rapidly appears as a result of global scientific research around the world.

It is especially important to develop visualization methods if we deal with analyzing unusual data, dynamically changing over time. In this case, visualization methods are faced with a very important and not easy task of visual and intuitive display of unrolling a certain value change over time, showing the nature of this change, analysis and cooperation with the accepted solutions, the result of these solutions. To do this, visualization types such as graphs, bar charts, and histograms are used.

However, the situation with visualization can become very complicated if the value in question is composite (consisting of several components). In certain cases, it is useful and very important for DM to analyze not just the change in the observed value, but also the impact of measures taken on the structure of its components, the dynamics of changes in the structure and the contribution of components. This paper is devoted to these data.

## 2 Problems of monitoring changes in structured values

Let us introduce the concept of a structured value, the visualization of dynamic changes in which we will consider in this paper. Let's suppose we have some observed value  $P$ . This value includes several components  $k_i$ :

$$P = \sum_i k_i. \quad (1)$$

Thus, the observed value is decomposed into a number of components. And from the point of view of visualizing such values, two cases can be distinguished:

- $P$  value is decomposed into a small number of components (less than 5);
- $P$  value consists of a large number of components.

Let's consider both cases using practical examples. During the quarantine period of the first half of 2020, a group of students and teachers at Bryansk State Technical University was engaged in constructing COVID-19 distribution models in AnyLogic [1]. At the same time, a number of modeled and observed values were just composite and had the character described above.

One of the most important and critical indicators of spreading COVID-19 is the number of infected people (let's call it INF). It is this indicator that primarily determines quarantine measures, the complexity of epidemic control in a particular region, and various management decisions [2,3]. However, as practice has shown, the total number of cases in a particular region (or even a country) does not adequately describe the situation. Even if this value is normalized by the region population (that is make the value relative), this is clearly not enough for the analysis. The fact is that INF value itself is a composite one by its nature and its constituent parts have quite different degrees of criticality and importance. For example, INF value can be decomposed into

such components as the number of seriously ill patients (requiring connection to an artificial lung ventilator), the number of moderate-severity patients, and the number of patients who do not require hospital admission. Therefore, the same INF value in different regions may show the overall situation in completely different ways. To understand the situation correctly, it is important to see not only the indicator itself, but also the structure of its components, and evaluate the contribution of each component.

Besides, if the number of components is not very large, we can use traditional schemes for displaying and visual analysis, such as multiple graphs (for each criterion separately), column graphs, histograms and Temporal Networks. In recent years, the research community has also accumulated overwhelming evidence in favor of complex and heterogeneous connectivity patterns based on Temporal networks [4]. However, the situation becomes more complicated if the number of components increases. In this case, the visibility of traditional approaches is greatly reduced.

Let's take another example. According to the World Health Organization (WHO), the incidence and severity of the disease in COVID-19 depends very much on the age. Therefore, it can be very useful to analyze the number of cases (INF indicator introduced before), but only in the context of the age of cases. That is, the components of this indicator should be considered as the number of cases of what age contributed to the total number of cases. This indication and analysis can be useful for evaluating actions and decisions taken. For example, some restrictive measures may not lead to a decrease in the total number of cases, but may change the structure of this value, reducing the number of elderly patients, which is also a very important factor in the epidemic control [5].

At the same time, visual analysis by age is very time-consuming. It is difficult to estimate this indicator due to the large number of age options. Therefore, in most cases, reduction of the variants (components) is done by rounding the age to certain ranges [6]. An example of such rounding, taken from official statistics on age, gender, and disease, depending on the death of patients from infection caused by COVID-19, based on an official document dated February 28, 2020 from the WHO-China Joint Mission report [7], is shown in Table 1.

**Table 1.** WHO-China Joint Mission data of depending the disease on the age

<b>Age</b>	<b>Number of infected (people)</b>	<b>Number of died (people)</b>	<b>Death probability due to COVID-19</b>
0-9 years	416	-	-
10-19 years	549	7	0,2%
20-29 years	3619	1	0,2%
30-39 years	7600	18	0,2%
40-49 years	8571	38	0,4%
50-59 years	10008	130	1,3%
60-69 years	8583	309	3,6%
70-79 years	3918	312	8,0%
80+ years	1408	208	14,8%

However, such rounding is associated with certain problems. Firstly, poorly chosen age ranges can lead to incorrect analysis and a strong distortion of the real situation. Secondly, by rounding up age ranges, we lose some of the information for analysis, which can be very significant. Thirdly, the selected ranges may vary greatly for different regions and countries, and to make the selection in each specific situation is a very time-consuming task [8]. Therefore, the important point is not to abandon the full range of ages, but to develop new principles for visualizing such data.

### 3 Existing render methods

There are a lot of ways to show the dynamics of data changes. Let us consider the most commonly used options by the example of dynamics of developing COVID-19 on the territory of Bryansk region (see Fig. 1). The graphs show the dynamics of infections, recoveries, deaths due to COVID-19, the dynamic number of new infections, identified on the territory of Bryansk region, compared with the previous day and according to days since the first revealed infected person.

The most common option is to use traditional charts. In this case, time is X-axis, and the value under consideration is Y-axis. If the data are structured, then several graphs are constructed for each of the components. An example of such graphs is shown in the first diagram of Fig. 1. However, it is good to analyze such graphs if the number of the observed values is small enough (one or two). If there are a lot of such values, the graph is difficult to read and it becomes very hard to draw any conclusions in this situation. Another problem is the difficulty of estimating the contribution of each component to the total. Based on such render methods, it is often impossible to analyze the overall situation of the epidemic growth in terms of the patient's age, gender, illness illnesses, activities, etc. [9].

Another option, which is also often found in open sources of statistics, is a column chart. In this case, each component is displayed as a separate column of its own color, and the dynamics is displayed as a set of such columns. According to Yandex, Apple and Otonomo data, there is given an example of a column chart grouping changes in the level of activity of the Russian population in the period from February to June.

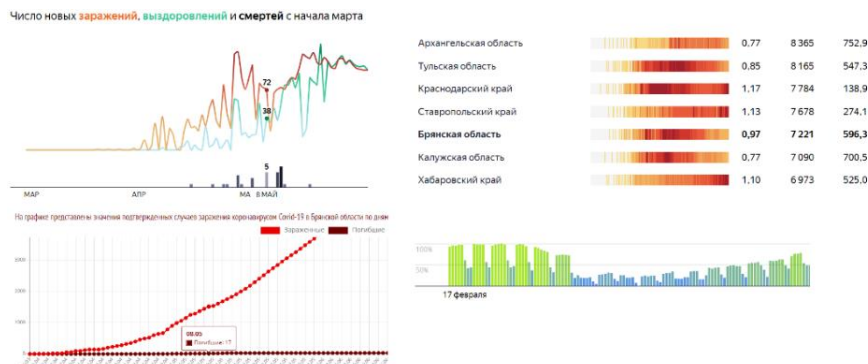


Fig. 1. Charts rendering the dynamics of COVID-19 in Bryansk region.

Another fairly good rendering option is a radar chart and a pie chart [10]. In contrast to the previous versions, these methods are very good at visualizing the data structure, clearly showing the contribution of each component. However, they are very difficult to apply for showing the dynamics of changes in this data over time.

#### 4 Method of rendering the graph of colour markers

As it was described above, in the framework of research and simulation of the situation with spreading COVID-19, conducted in Bryansk State Technical University, we used an approach to show the dynamics of changes in structured values, which is conventionally called the method of color stripes or color marker graph. The main idea of this approach is as follows. Let us assume that we have a certain value  $P(t)$ , which value changes over time (the value under study depends on  $t$  parameter). In this case, this value consists of different components. That means that at  $t$  time moment  $P(t)$  value can be decomposed into some components:

$$P(t) = \sum_i k_i(t). \quad (2)$$

For visual image of these data (both the graph of changes in the value itself and its components), we assign each component part  $k_j$  a certain color (the so-called color marker):

$$\{k_1 \leftrightarrow \text{цвет}_1; k_2 \leftrightarrow \text{цвет}_2; \dots; k_n \leftrightarrow \text{цвет}_n\}. \quad (3)$$

Then the curve of the value change  $P(t)$  is rendered as a standard graph, plotting time along the abscissa axis, and the observed value itself along the ordinate axis. At the same time,  $k_j(t)$  component parts should be painted with the corresponding color marker. In this case, the graph of changes  $P(t)$  will look like color bars of different widths (depending on the value of  $k_j(t)$  components).

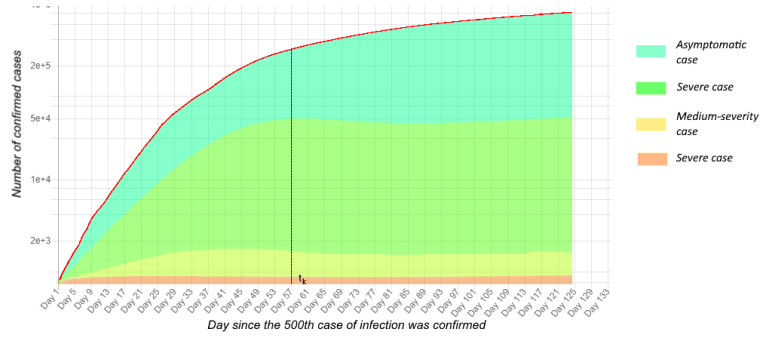
Now let us consider the example described in the previous chapter (analysis of changes in the number of cases in the context of disease severity). Let the analyzed value  $P(t)$  be the total number of infected by COVID-19 at a certain time. At the same time, as it is described above, not the graph of changes in the number of cases is of interest, but the analysis of the components of this number, that is how many of these cases are severe, how many are moderate, how many people do not require hospital admission, and how many are asymptomatic. In this case this value at time moment  $t$  is decomposed as follows:

$$INF(t) = \sum_{i=1}^4 k_i(t). \quad (4)$$

where:  $k_1(t)$  is the number of seriously ill at time moment  $t$ ;  $k_2(t)$  is the number of hospital moderate-severity patients;  $k_3(t)$  is the number of infected people who do not need hospital admission;  $k_4(t)$  is the number of asymptotically infected.

Let us assign a color marker to each value. Then the general graph of changes in the epidemiological situation will take the form shown in Fig. 2.

This graph is a more visual way to show than tabular data or other options discussed earlier. This display provides more information for analysis and additional conclusions, allows to see patterns or effects that are not visible in other versions of the image. For example, according to the graph above, it can be seen that at time moment  $t_k$  marked on the graph with a stipple line, some measures were taken that did not significantly reduce the increase in the total number of patients, but as a result, they significantly changed the structure of this value in the direction of reducing the number of seriously ill and moderate-severity patients.



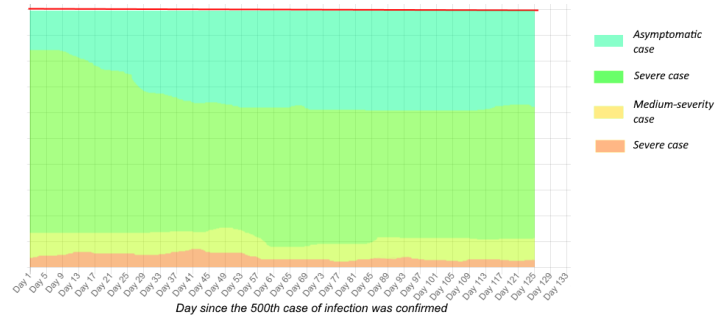
**Fig. 2.** Graph for showing a structured value in the form of color markers in the context of disease severity.

Sometimes searching for such structural changes is even more important than analyzing the total value. In this case, we can switch from absolute numbers to relative ones, i.e. construct color bars in a normalized variation:

$$k'_i(t) = k_i/P(t), \quad (5)$$

$$\sum_{i=1}^n k'_i(t) = 1. \quad (6)$$

In this case, the analysis of changes in the structure becomes even more clear and any changes or external influences that displace the internal components of the value become clearer (see Fig. 3).



**Fig. 3.** Graph of structural changes.

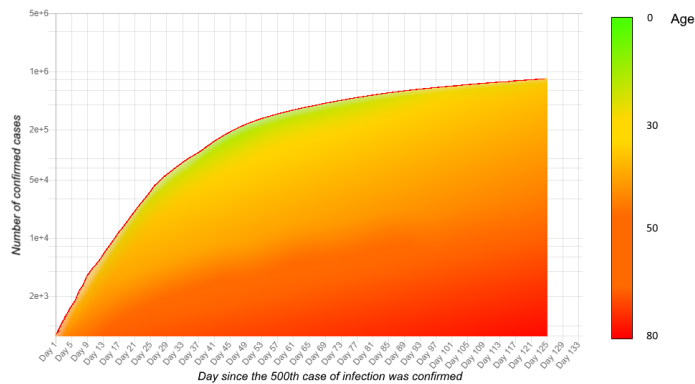
Let us consider another example. In this case the observed value  $P(t)$  consists of a large number of components. An example of such data is the previously considered analysis of morbidity rate in the context of patients' age. As practice has shown, this parameter is also very important for analyzing the situation.

Formally, this task is expressed as follows. As before, we will use the number of patients at time moment  $t$  as the observed value. However, we will consider the value of the composite value  $k_{age}(t)$  as the number of cases at time  $t$  at the age, where  $age$  is in the range:

$$0 \leq age \leq MaxAge, \tag{7}$$

where  $MaxAge$  is the maximum age of infected people.

Visual representation of such a value by conventional methods (graphs, charts, tables) is quite problematic [11,12]. In the proposed approach, in order to use color markers, we should enter a smoothly changing color marker for ages. In this case, the graph shown in Fig. 4 will be obtained.



**Fig. 4.** Graph of showing structured value by means of color markers in the context of patients' age.

This display does not require analysis and selection of rounding limits, but it is a simpler and more visual way to display a composite value.

## 5 Data visualization in the software system for modeling the situation with COVIND-19 in Bryansk region

Epidemic outbreaks in various regions represent a special situation with a high level of uncertainty. In some regions, there was a noticeable repetition of the situation in the same scenario. In others, the situation was significantly different. A software package in AnyLogic environment was developed in Bryansk State Technical University. Classical methods of epidemic modeling based on SIR model were suggested as a basis [13]. As a result of analyzing a series of experiments conducted with COVID-19 disease development model it is possible to visualize the rate of infection spread with different

behavior of people and various parameters, conditions and constrains. The screenshot of the system is shown in Fig. 5. The system was presented at the contest for the best scientific work "Modern scientific achievements. Bryansk-2020".

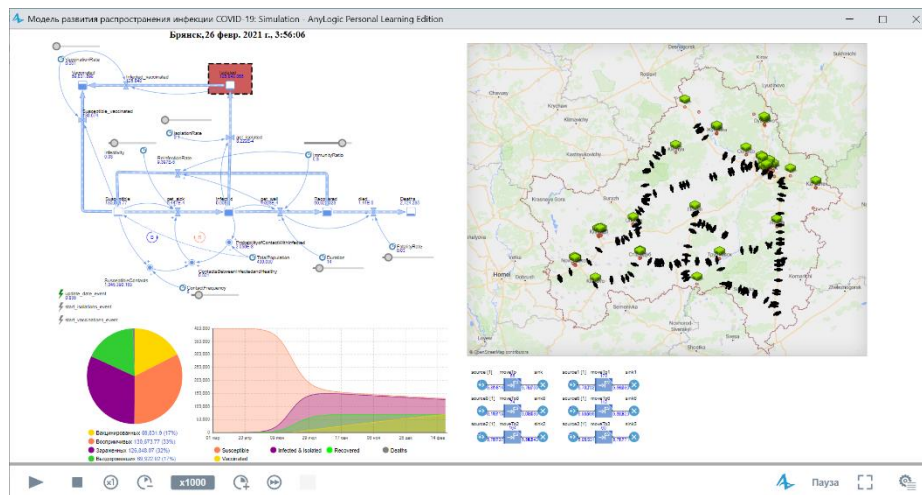


Fig. 5. System of modelling COVID-19 in Bryansk region.

Taking the known criteria as a basis, it is possible to demonstrate the development of infection spread, indicators of the level of disease severity [14]. We used a model that incorporated the main factors known at that time about the spread of infection among people. Specific indicators and coefficients were calculated based on similar data for Moscow region [15] and included in the complex for modeling the situation in Bryansk region. As a result, a model was built (see Fig. 6) that takes into account the following factors:

1. Immunity. The better the person's immunity, the lower is the probability of COVID-19 infection (*ImmunityRatio* parameter of double type, which takes value from 0 to 1). Each object *Person* in the constructor is assigned with a random value in the range specified above.
2. Social responsibility (*IsolationRate* parameter of double type has a default value of 0.45) [16]. According to the emergency response centre of the Russian Federation, the level of social responsibility is approximately 45% (exactly following the doctor's prescription and recommendations, as well as quarantine or self-isolation measures). The value of this parameter changes dynamically based on data received from Rospotrebnadzor and Johns Hopkins University.
3. Accompanying coronavirus diseases (factor of a person's going to the hospital in case of infection, *accompanyingIllness* parameter of boolean type).
4. The time to make a decision after infection symptoms onset is a factor that can be used to make a conclusion about the disease severity (*whatToDoDays* parameter is of int type and is set randomly from the above range for each object *Person*). According to statistics in the region [17], a person goes to the hospital or goes to self-



isolation within 1-5 days. In this time range, a person begins to notice symptoms onset and deterioration of his health condition.

The structure of the simulation model of the disease development according to the unit of model time equal to one day is shown in Fig. 6.

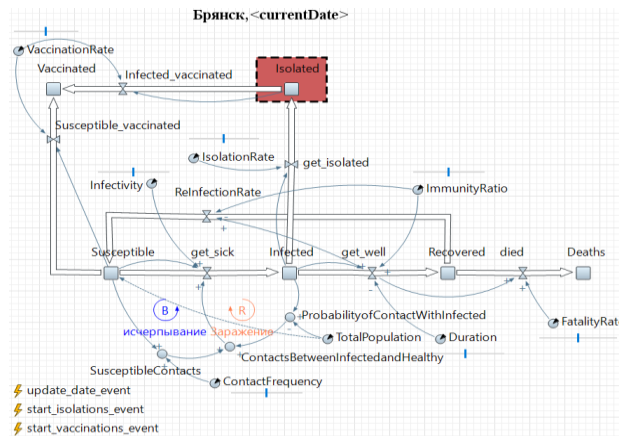


Fig. 6. Logical model of the system.

The color marker approach described above was used to display and visualize data on the dynamics of changes in structured values. This made it possible to present the results more clearly.

The data on the infection spread in Bryansk region obtained during modeling on the basis of open and publicly available information shows the feasibility of the model. The discrepancy between the currently modeled and real curves does not exceed 12%.

## 6 Conclusion

In this paper, we have proposed a method that combines analysis and visualization of structured data. Most of the examples in the paper are given on the situation with COVID-19 development. This is due to the fact that we have tried and tested this visualization approach on the basis of the corresponding complex. However, the proposed method of displaying the dynamics of changes in the data structure in the form of color markers and gradients can be used in absolutely any areas where there is a need to display the dynamics of composite values, see and analyze the structure of these values.

## References

1. AnyLogic, <https://www.anylogic.ru/>, last accessed 2020/07/10.
2. Yang, L.Y., Yan, L.M., Wan, L., Xiang, T.-X., Le, A., Liu, J.M., Peiris, M., Poon, L.L.M., Zhang, W.: Viral dynamics in mild and severe cases of Covid 19. *Lancet Infect Dis.* (2020).

3. Murray, C.J.L.: Forecasting COVID 19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. IHME COVID 19 health service utilization forecasting team, <https://doi.org/10.1101/2020.03.27.20043752>, last accessed 2020/07/10.
4. Kapoor, A, Ben, X, Liu, L, et al.: Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks. July 06, 2020 (2020).
5. Koltsova, E.M., Kurkina, E.S., Vasetsky, A.M.: Mathematical modeling of the spread of COVID-19 coronavirus epidemic in a number of European, Asian countries, Israel and Russia. *Economic problems and legal practice*, no.2 (2020).
6. Russell Timothy, W , Hellewell, J., Jarvis Christopher, I., van Zandvoort, K., Abbott, S., Ratnayake, R., Flasche S., Eggo, R.M., Edmunds, W.J., Kucharski, A.J.: Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020, <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256>, last accessed 2020/07/10.
7. World Health Organization. Coronavirus disease 2019 (COVID-19): Situation Report – 38 from 27 February 2020, <http://www.who.int/docs/default-source/coronaviruse/situation-reports/20200227-sitrep-38-covid-19.pdf>, last accessed 2020/07/10.
8. Zhang, S-H, Cai, Y, Li J.: Visualization of COVID-19 spread based on spread and extinction indexes. *Sci China Inf Sci*, 63 (2020).
9. Yang, X, Yu, Y, Xu, J, et al.: Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 8(5), pp. 475-481 (2020).
10. Dimara, E, Perin, C.: What is interaction for data visualization? *IEEE Trans Visual Comput Graph* 26, pp. 119-129 (2020).
11. Robertson, G, Fernandez, R, Fisher, D, et al.: Effectiveness of animation in trend visualization. *IEEE Trans Visual Comput Graph* 14, pp. 1325-1332 (2008).
12. Kosara, R.: Presentation-oriented visualization techniques. *IEEE Comput Grap Appl*, 2016, 36, pp. 80-85 (2016).
13. Khrapov, P.V., Loginova, A.A.: Mathematical modelling of the dynamics of the coronavirus COVID-19 epidemic development in China. *International Journal of Open Information Technologies* 8(4), pp. 13-16, <http://www.injoit.org/index.php/j1/article/view/908/874>, last accessed 2020/07/10.
14. Wang, D.Q, Guo, D.H, Zhang, H.: Spatial temporal data visualization in emergency management: a view from data-driven decision. In: *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Emergency Management*, pp. 1-7 (2017).
15. Matveev, A.V.: Mathematical modeling of evaluating the effectiveness of measures against the spread of COVID-19. *Natsionalnaya Bezopasnost I Strategicheskoe Planirovanie (National Security and Strategic Planning)*, no.1(29), pp. 23-39 (2020).
16. Jüni, P., Rothenbühler, M., Bobos, P., Thorpe, K.E., da Costa, B., Fisman, D., Slutsky, A.S., Gesink, D.: Impact of climate and public health interventions on the COVID-19 pandemic: A prospective cohort study. *CMAJ* May 08, 2020 (2020).
17. Rodkin, M.V., Shikhova, N.M: Mathematical modeling of COVID-19 epidemic, an attempt to forecast. *Uralian Geological Journal*, no.3 (2020).