Spatial Distribution of COVID-19 Under the Influence of Environmental Factors: Correlation and Regression Analysis

N V Kostina, R S Kuznetsova and G S Rozenberg
Samara Federal Research Scientific Center RAS, Institute of Ecology of Volga River Basin RAS, Togliatti, Russia

E-mail: knva2009@yandex.ru

Abstract. Close monitoring of morbidity and mortality from COVID-19 has become a "supplier" of large amounts of spatially distributed information. This, in turn, has led to the creation of many models for the development of a pandemic. We used official information on the incidence and mortality of COVID-19 for 24 regions (oblasts and republics) of the Volga basin of the Russian Federation. We performed multiple regression analysis to test the significance of the impact of 19 social, economic and environmental indicators. 10 parameters turned out to be statistically significant for the number of cases, among which social indicators make a significant contribution (63.19%). For the indicator "mortality from COVID-19", we identified 15 statistically significant factors. At the same time, we note the prevalence of socio-economic parameters (58.4%) over the environmental ones (31.2%). Among the analyzed 19 indicators, there were no such indicators with the help of which it is possible to quickly influence the incidence and mortality from COVID-19. Vaccination of the population is still a topical remedy. Nevertheless, we consider the very approach to the study of spatially distributed information to be successful. By combining the efforts of infectious disease specialists, systemologists, ecologists and other specialists, one can hope to create a workable (predictive and manageable) model for the development of COVID-19.

1. Introduction
2020 was marked by COVID-19, a new coronavirus infection. As of December 31, 2020, the pandemic had recorded over 83.2 million cases worldwide; more than 1.815 million people died and more than 59.29 million - recovered [1]. In this article, we will not dwell on the features of the etiology, transmission of infection, pathogenesis, symptoms, etc. of this disease - all this is widely covered in the media and in special sources. We will only emphasize that monitoring of morbidity and mortality from COVID-19 has become a "supplier" of large amounts of spatially distributed information. This, in turn, has led to the creation of many models for the development of a pandemic [2-6, etc.]. However, the creation of predictive models, even in the presence of large amounts of information, is a rather complicated and ambiguous business in terms of the result obtained. Thus, according to the morbidity model created by Sberbank [7], if everyone observes safety measures, the peak will be passed on November 8, and the number of new cases per day will be 16.8 thousand people. (On November 8, there were 20.5 thousand cases and the growth continued [the next 10 days on average for 21.7 thousand people]). The peak of the incidence fell on December 24 and amounted to 29.9 thousand people per day…
A more accurate forecast is provided by a mathematical model of the spread of coronavirus infection, which was developed by employees of the Competence Center of the National Technological Initiative on the basis of the Peter the Great St. Petersburg Polytechnic University together with experts from the Institute of Biomedical Systems and Biotechnology of the same university and the Research Institute of Influenza. The project manager for the creation of a mathematical model is Vice-Rector for Advanced Projects of the University, Associate Professor A.I. Borovkov. Based on the simulation results, back in the autumn, Borovkov made a forecast: "The peak of the pandemic in the capital will be in mid-December," and then "at the end of January 2021, there will be about 115 thousand active patients in Moscow - the same number as on May 19 at the first peak. " [8]. According to the CORONAVIRUS (COVID-19) website, as of January 29, 2021, there were 108,416 patients in Moscow [9]; forecast error is less than 6%, moreover, in the direction of "underestimating" the result.

2. Methods and Materials

To conduct our research, we used data from open sources, the official information provided on the CORONAVIRUS website (COVID-19), as well as the environmental information system REGION [10], developed at the Institute of Ecology of the Volga River Basin of the Russian Academy of Sciences. The database contains social, economic and environmental spatially distributed information for 24 regions (oblasts and republics) located on the territory of the Volga basin. In the study, we used the method of correlation-regression analysis with checking the significance of the influence of the factors under study by the method of I Ya Liepa [11, 12].

3. Results and Discussion

Distribution of cases across the Volga basin as of August 31, 2020 is shown in figure 1. The minimum value of the number of cases per 1,000 people is 1.67 (The Republic of Tatarstan), maximum value is 9.82 (Ulyanovsk Oblast). Nizhny Novgorod, Moscow and Kaluga oblasts are among the territories with the maximum incidence on this date.

The number of deaths from COVID-19 as of August 31, 2020 is shown in figure 2. The maximum value (0.22 cases per thousand people) was in Tula Oblast.

Figure 1. Distribution of the number of COVID-19 cases over the territory of the Volga basin as of August 31, 2020 (number of cases per 1,000 people)

1 – from 1.67 to 3.71;
2 – from 3.71 to 5.74;
3 – 5.74 to 7.78;
4 – more than 7.78.
Using the environmental information system REGION, we conducted multiple regression analysis, estimated the parameters of the multiple linear regression equation using the least squares method and checked the significance of the influence of the studied factors for the number of COVID-19 cases. We took 19 socio-ecological and economic indicators into consideration [13]:

- GAD - density of hard-surface roads (km / 1000 km$^2$),
- SZSV - discharge of polluted wastewater into surface water bodies (million m$^3$ / thousand people),
- ZVA - emissions of contaminated substances into the atmosphere from stationary sources (t / person),
- LT - forest cover of the territory (%),
- EF - ecological footprint (gha per person),
- HDI - Human Development Index,
- KG - the Gini coefficient,
- KZ is an integral indicator of the quality of life,
- DD is the proportion of children (under 16),
- DS is the proportion of the elderly (men over 60 and women over 55),
- HE - percentage of people with higher education,
- KR - fertility rate,
- Svos - average age as of January 1, 2019,
- AIPC - average per capita income (in rubles),
- SNE - share of natural ecosystems area (%),
- KVZA - number of emissions from stationary sources, kg / ha / year,
- UFW - fresh water use, m$^3$ / ha / year,
- VZS - volume of polluted runoff, m$^3$ / ha / year,
- ND - number of doctors, per 10 thousand people.

Ten parameters turned out to be statistically significant (table 1).

Environmental parameters KVZA, UFW, GAD, HDI (partly), KZ (partly) are significant in this equation. The sum of the specific influence of these factors is 7.94%. The contribution of the social component is more significant - 63.19, and if we take into account that the HDI and KZ also contain socio-economic indicators, then the share of the latter will increase even more.
Table 1. Multiple regression equation for estimating Y (number of cases of COVID-19) from statistically significant socio-ecological-economic indicators

| Factors included in the regression model | Regression coefficients | The share of the influence of factors (%) |
|-----------------------------------------|-------------------------|------------------------------------------|
| Free term                               | 157,77                  |                                          |
| GAD                                     | -0,0046                 | 1,44                                     |
| HDI                                     | -49,495                 | 3,00                                     |
| KZ                                      | 0,2605                  | 1,33                                     |
| DD                                      | -3,68                   | 35,92                                    |
| DS                                      | -1,11                   | 11,22                                    |
| KR                                      | 11,318                  | 8,42                                     |
| Svos                                    | -0,935                  | 6,59                                     |
| KVZА                                    | -0,038                  | 1,43                                     |
| UFW                                     | 0,0034                  | 0,74                                     |
| ND                                      | -0,22                   | 1,04                                     |

Accumulated sum of specific influence of factors 71,13

Multiple correlation coefficient 0,843

The multiple regression equation for the COVID-19 mortality rate contains 15 statistically significant factor-parameters (table 2).

Table 2. Multiple regression equation for estimating Y (number of deaths from COVID-19) from statistically significant socio-ecological-economic indicators

| Factors included in the regression model | Regression coefficients | The share of the influence of factors (%) |
|-----------------------------------------|-------------------------|------------------------------------------|
| Free term                               | 5,295                   |                                          |
| GAD                                     | -0,0002                 | 2,78                                     |
| VZVA                                    | -0,109                  | 0,48                                     |
| LT                                      | -0,0012                 | 4,40                                     |
| HDI                                     | -2,815                  | 11,55                                    |
| KG                                      | 0,350                   | 0,68                                     |
| KZ                                      | -0,012                  | 1,88                                     |
| DD                                      | -0,062                  | 22,51                                    |
| DS                                      | 0,014                   | 7,29                                     |
| HE                                      | -0,0058                 | 1,26                                     |
| KR                                      | 0,198                   | 3,33                                     |
| Svos                                    | -0,068                  | 20,34                                    |
| SNE                                     | 0,0018                  | 7,35                                     |
| UFW                                     | 0,00017                 | 3,72                                     |
| VZS                                     | -0,00002                | 0,94                                     |
| ND                                      | 0,0161                  | 1,09                                     |

Accumulated sum of specific influence of factors 89,60

Multiple correlation coefficient 0,947

When comparing tables 1 and 2, we note that all factors-parameters of morbidity from COVID-19 were included in the equation for mortality from COVID-19, except for KVZА (the number of emissions from stationary sources). It should also be noted that in both cases socio-economic parameters prevail over the environmental ones (the contribution of environmental parameters -
VZVA, LT, KVZA, UFW, GAD, SNE, VZS, HDI, KZ, KG [the last three parameters - partially] – 31.2, socio-economic - 58.4, with the same remark that the HDI, KG and KZ also contain socio-economic indicators), although the difference is no longer so great.

We can draw one more conclusion after the conducted research. Among the analyzed 19 indicators, there were no such indicators with the help of which it is possible to quickly influence Y. However, the very approach to studying spatially distributed information on morbidity and mortality from COVID-19 using the EIS REGION should be considered successful. By combining the efforts of infectious disease specialists, systemologists, ecologists and other specialists, one can hope to create a workable (predictive and manageable) model for the development of COVID-19.

4. Conclusion
Since 2020, we have been living in a new reality. And, most likely, it will soon be necessary to add new goals to the 17 main Sustainable Development Goals and set them very high in priority. In particular, we will take the liberty of proposing the following options: "To unite efforts, to conduct constant monitoring and to be prepared for the emergence of pandemics of different natures», some sort of natural-spiritual fast.

Funding
This work was carried out within the framework of the Program of Fundamental Research of the State Academies of Sciences in 2013-2020 (projects nos. AAAA-A17-117112040040-3 and AAAA-A17-117112040039-7).

References
[1] The Covid-19 pandemic URL: https://ru.wikipedia.org/wiki/Пандемия_COVID-19
[2] Koltsova E M, Kurkina E S and Vasetsky A M 2020 Mathematical modeling of the spread of the COVID-19 epidemic in Moscow Computational Nanotechnology 7(1) 99-105
[3] Sukhovolsky V G and Kovalev A V 2020 Modeling the Coronavirus Epidemic as a Phase Transition J. of General Biology 81(5) 362-373
[4] Cherniha R., Davydovych V. 2020 A mathematical model for the coronavirus COVID-19 outbreak // arXiv preprint arXiv: 2004.01487.
[5] Qi C, Karlsson D, Sallmen K and Wyss R 2020 Model studies on the COVID-19 pandemic in Sweden // arXiv preprint arXiv: 2004.01575.
[6] Sameni R. 2020 Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus. arXiv preprint arXiv 2003.11371.
[7] Sberbank predicted the peak incidence of coronavirus in Russia URL: https://tass.ru/obschestvo/9672345?utm_source=vyxnews&utm_medium=desktop
[8] The peak of the pandemic in the capital will be in mid-December… URL: https://news.myseldon.com/ru/news/index/240124632
[9] Coronavirus statistics in Moscow on January 29, 2021 URL: https://coronavirus-monitorus.ru/v-rossii/koronavirus-v-moskve-situatsiya-na-29-yanvarya-2021/
[10] Kostina N V 2015 Information system REGION: 25 years of development and practical application Bulletin of Samara Scientific Center of RAS 17(4) 15-24
[11] Liepa I Ya 1971 Indicator of the specific influence of influencing factors Scientific Notes of the Latvian University (Riga: Latvian university) pp 36-40
[12] Liepa I Ya 1978 Typical errors in the application of mathematical methods in biology Modeling and forecasting in ecology (Riga: Latvian university) pp 3-14
[13] Bukvareva E at al 2016 Ecosystem services of Russia. Prototype of national report. Services of terrestrial ecosystems 1 (Moscow: Wildlife Conservation Center Publishing)