Long-term predictions for COVID-19 pandemic dynamics in Ukraine, Austria and Italy

Igor Nesteruk

Institute of Hydromechanics. National Academy of Sciences of Ukraine.
Zheliabova St. 8/4. UA-03680 Kyiv. Ukraine

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".
Peremohy Av. 37. UA-03056. Kyiv. Ukraine
inesteruk@yahoo.com

ABSTRACT

The SIR (susceptible-infected-removed) model, statistical approach to the parameter identification and the official WHO daily data about the confirmed cumulative number of cases were used to make some estimations for the dynamics of the coronavirus pandemic dynamics in Ukraine, Italy and Austria. The volume of the data sets and the influence of the information about the initial stages of the epidemics were discussed in order to have reliable long-time predictions. The final sizes and durations for the pandemic in these countries are estimated.

Keywords: coronavirus pandemic, epidemic outbreak in Italy, Austria and Ukraine, coronavirus COVID-19, coronavirus 2019-nCoV, mathematical modeling of infection diseases, SIR model, parameter identification, statistical methods.

Introduction

Here we consider the development of epidemic outbreak in Italy, Ukraine and Austria caused by coronavirus COVID-19 (2019-nCoV) (see e.g., [1]). Some estimations of the epidemic dynamics in these countries can be found in [2-7]. In particular, the final size of the epidemic in Italy was calculated in [5]. Unfortunately, the real number of confirmed cases in Italy already exceeds the saturation level predicted in [5] on March 27, 2020. In this paper we will try to clarify the predictions for Italy, to estimate the final size and the duration of the epidemic in Austria and Ukraine and to discuss the amounts of data needed for accurate long-term predictions. In this paper we will use the SIR model [8–12] and the statistics-based method of parameter identification [11, 12].

Data

The official data about the accumulated numbers of confirmed COVID-19 cases in Italy V_j , Ukraine V_{Uj} and Austria V_{Aj} from WHO daily situation reports (numbers 41-78), [1] are presented in Tables 1 and 2. The corresponding moments of time t_j (measured in days) are also shown in these tables. The data sets presented in Table 1 were used only for comparison with corresponding SIR curves. Table 2 was used for calculations, comparisons and verifications of predictions.

Day in February and March, 2020	Time moments In days t_j	Number of cases in Italy, V_J	Number of cases in Austria, V_{Aj}	Number of cases in Ukraine, V_{Uj}
21	-1	9	0	0
22	0	76	0	0
23	1	124	0	0
24	2	229	0	0
25	3	322	2	0
26	4	400	2	0
27	5	650	4	0
28	6	888	5	0
29	7	1128	10	0
1	8	1689	10	0
2	9	2036	18	0
3	10	2502	24	1
4	11	3089	37	1

Table 1. Official cumulative numbers of confirmed cases in Italy, Austria and Ukraine during the initial stage of epidemics used only for comparison with SIR curves, [1]

SIR model

The SIR model for an infectious disease [7-11] relates the number of susceptible persons S (persons who are sensitive to the pathogen and **not protected**); the number of infected is I (persons who are sick and **spread the infection**; please don't confuse with the number of still ill persons, so known active cases) and the number of removed R (persons who **no longer spread the infection**; this number is the sum of isolated, recovered, dead, and infected people who left the region); α and ρ are constants.

$$\frac{dS}{dt} = -\alpha SI \tag{1}$$

$$\frac{dI}{dt} = \alpha SI - \rho I \tag{2}$$

$$\frac{dR}{dt} = \rho I \tag{3}$$

To determine the initial conditions for the set of equations (1–3), let us suppose that at the moment of the epidemic outbreak t_0, [10, 11]:

$$I(t_0) = 1, \; R(t_0) = 0, \; S(t_0) = N - 1, \; N = S + I + R \tag{4}$$

Day in March and April, 2020	Time moments In days t_j	Number of cases in Italy, V_J	Number of cases in Austria, V_{Aj}	Number of cases in Ukraine, V_{Uj}
5	12	3858	47	1
6	13	4636	66	1
7	14	5883	104	1
8	15	7375	112	1
9	16	9172	131	1
10	17	10149	182	1
11	18	12462	302	1
12	19	15113	361	3
13	20	17660	504	3
14	21	21157	800	3
15	22	24747	959	3
16	23	27980	1132	7
17	24	31506	1332	14
18	25	35713	1646	16
19	26	41035	1843	16
20	27	47021	2649	26
21	28	53578	3024	47
22	29	59138	3631	47
23	30	63927	4486	84
24	31	69176	5282	113
25	32	74386	5888	156
26	33	80539	7029	218
27	34	86498	7697	311
28	35	92472	8291	418
29	36	97689	8813	480
30	37	101739	9618	549
31	38	105792	10182	669
1	39	110574	10711	804
2	40	115242	11129	987
3	41	119827	11525	1096
4	42	124632	11766	1251
5	43	128948	11983	1319
6	44	132547	12297	1462

Table 2. Official cumulative numbers of confirmed cases in Italy, Austria and Ukraine used for calculations, comparisons and verifications of predictions, [1]

The analytical solution for the set of equations (1–3) was obtained by introducing the function $V(t)=I(t)+R(t)$, corresponding to the number of victims or cumulative confirmed number of cases, [11]:

$$F_1(V,N,\nu)=\alpha(t-t_0) \qquad (5)$$

$$F_1=\int_1^V \frac{dU}{(N-U)\left[\nu\ln(N-U)+U-\nu\ln(N-1)\right]},\qquad \nu=\frac{\rho}{\alpha} \tag{6}$$

Thus, for every set of parameters *N*, ν, α, t_0 and a fixed value of V the integral (6) can be calculated and the corresponding moment of time can be determined from (5). Then functions *I(t)* and *R(t)* can be easily calculated with the of formulas, [11, 12].

$$I=\nu\ln S-S+N-\nu\ln(N-1),\qquad S=N\text{-}V,\qquad R=V\text{-}I. \tag{7}$$

Function I has a maximum at $S=\nu$ and tends to zero at infinity, see [8, 9]. In comparison, the number of susceptible persons at infinity $S_\infty>0$, and can be calculated from the non-linear equation, [11, 12]:

$$S_\infty=(N-1)e^{\frac{S_\infty-N}{\nu}} \tag{8}$$

The final number of victims (final accumulated number of cases) can be calculated from:

$$V_\infty=N-S_\infty \tag{9}$$

To estimate the duration of an epidemic outbreak, we can use the condition

$$V(t_{final})=1 \tag{10}$$

which means that at $t>t_{final}$ less than one person still spread the infection.

Parameter identification procedure

In the case of a new epidemic, the values of this independent four parameters are unknown and must be identified with the use of limited data sets. A statistical approach was developed in [11] and used in [5, 11, 12, 15] to estimate the values of unknown parameters. The registered points for the number of victims V_j corresponding to the moments of time t_j can be used in order to calculate $F_{1j}=F_1(V_j,N,\nu)$ for every fixed values N and ν with the use of (6) and then to check how the registered points fit the straight line (5). For this purpose the linear regression can be used, e.g., [13], and the optimal straight line, minimizing the sum of squared distances between registered and theoretical points, can be defined. Thus we can find the optimal values of α, t_0 and calculate the correlation coefficient r.

Then the F-test may be applied to check how the null hypothesis that says that the proposed linear relationship (5) fits the data set. The experimental value of the Fisher function can be calculated with the use of the formula:

$$F = \frac{r^2(n-m)}{(1-r^2)(m-1)} \tag{11}$$

where n is the number of observations, m=2 is the number of parameters in the regression equation, [13]. The corresponding experimental value F has to be compared with the critical value $F_C(k_1, k_2)$ of the Fisher function at a desired significance or confidence level ($k_1 = m-1,\ k_2 = n-m$), [14] . When the values n and m are fixed, the maximum of the Fisher function coincides with the maximum of the correlation coefficient. Therefore, to find the optimal values of parameters N and ν, we have to find the maximum of the correlation coefficient. To compare the reliability of different predictions (with different values of n) it is useful to use the ratio $F/F_C(1, n-2)$ at fixed significance level, [15]. We will use the level 0.001; corresponding values $F_C(1, n-2)$ can be taken from [14]. The most reliable prediction yields the highest $F/F_C(1, n-2)$ ratio.

Results for Italy

The first preliminary prediction for Italy was published in [5] on March 27, 2020. Its results are presented in the first column of Table 3. Usually the number of cases during the initial period of an epidemic outbreak is not reliable. To avoid their influence on the results, only V_j values for the period March 5-22, 2020 ($12 \le t_j \le 29$, see Table 2) were used to calculate this first prediction. For other predictions different values of the initial t_{j1} and final t_{j2} moments of time were used. They are shown in Tables 3-5 together with corresponding number of observations n. The SIR curves for the prediction No. 4 (calculated with the use of the recent V_j values without taking into account maximum number of initial moments of time) are shown in Fig. 1. Markers represent the V_j values taken for calculations (“circles”); for a comparison (“triangles”) and a verification of predictions (“star”).

It can be seen that the real epidemic outbreak could start even in January 2020. Probably, first cases of COVID-2019 infection were not identified and sick people were not isolated. It could be a reason of the high saturation level of the epidemic (approximately 148,000 - 153,000 according to the last predictions 3 and 4). The new cases could stop to appear at the moments of time 93 -99 (see two last values in the last row of Table 3). These moments correspond to May 25-31, 2020. The average time of spreading infection $1/\rho$ could be estimated as 0.7-0.75 days (according to last two predictions, see Table 3). By comparison, in South Korea was approximately 4.3 hours, [16].

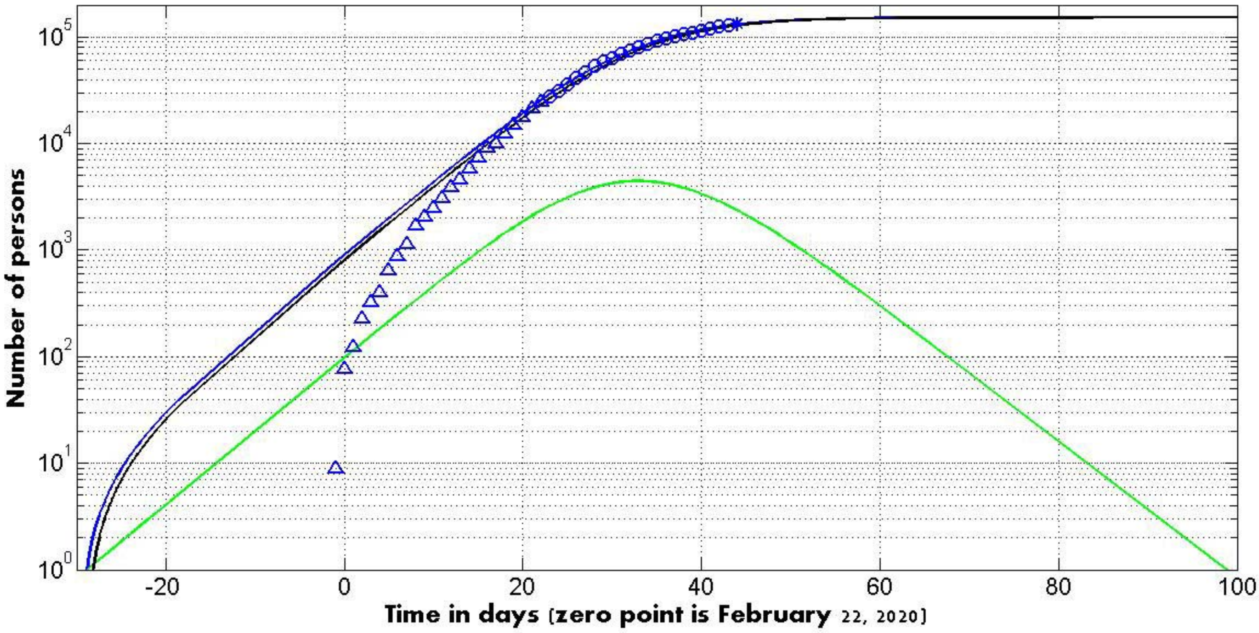

Fig. 1. Italy: SIR curves (lines) and accumulated number of cases (markers) versus time

Numbers of infected I (green), removed R (black) and the number of victims $V=I+R$ (blue line).

Number of prediction	**1, [5], March 27, 2020**	**2**	**3**	**4**
N	534656	669312	723763.2	738240
ν	476712.7441113	598876.49226424	646709.93100256	658709.4638592
α	3.47467962e-06	2.589545071e-06	2.2076760085e-06	2.0049693740e-06
t_0	-18.3797687613	-22.039683744	-25.171859952631	-28.810667619253
ρ	1.656424058815	1.5508176688225	1.4277259991363	1.3206923014088
$1/\rho$	0.603710139730	0.6448211289462	0.7004145057279	0.7571786395160
t_{j1}	12	17	17	23
t_{j2}	29	37	43	43
n	18	21	27	21
r	0.999378228510	0.999725363341	0.9992432785563	0.9992024895
F, eq. (11)	12854.46446586	34576.908427910	16499.880328838	11897.820763752
$F_C(1,n-2)$	16.2	15.2	13.9	15.2
$F/F_C(1,n-2)$	793.4854608556	2274.7966070993	1187.0417502761	782.75136603635
S_∞, eq (8)	423108.1863318	533557.58021716	575322.57589364	585099.75620840
V_∞, eq (9)	111547.8136681	135754.419782840	148440.624106361	153140.243791598
t_{final}, eq (10)	79.5	87.8	93.9	98.8

Table 3. Predictions for the epidemic in Italy. Optimal values of parameters and other SIR model characteristics.

Results for Austria

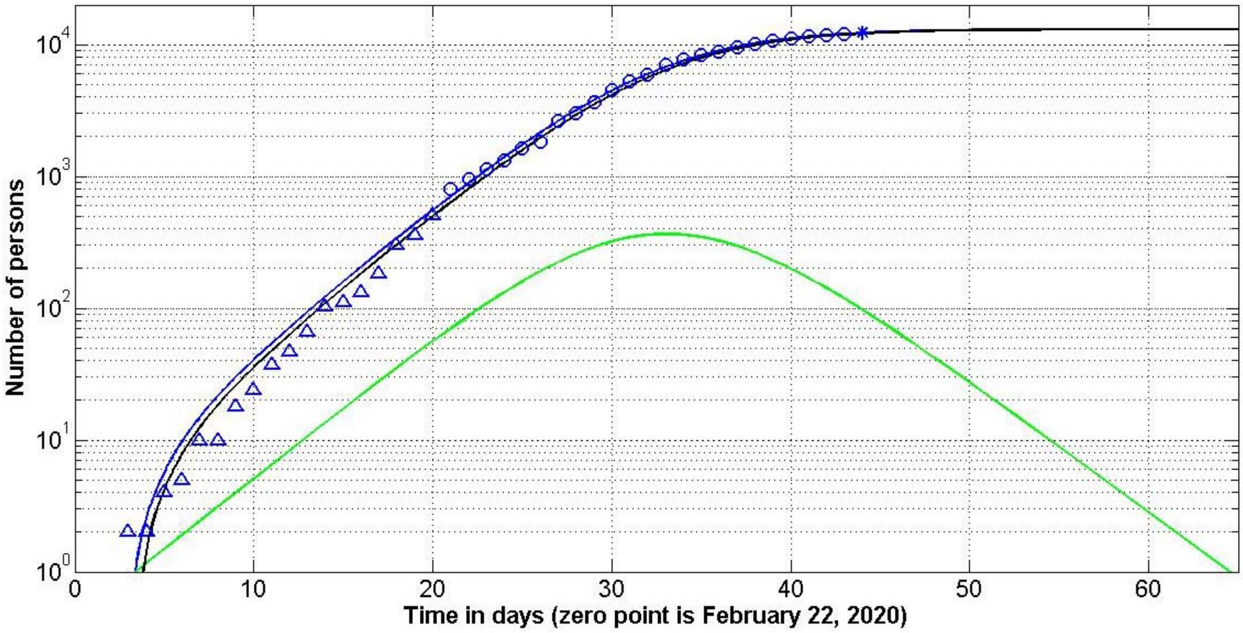

Fig. 2. Austria: SIR curves (lines) and accumulated number of cases (markers) versus time

Numbers of infected I (green), removed R (black) and the number of victims $V=I+R$ (blue line); “circles” show the cases taken for calculations; “triangles” correspond to the cases during initial stage of the epidemic; “star” –the last data point used only for a verification of the prediction.

Number of prediction	**1**	**2**	**3**
N	71099.2204800000	65613.6960000000	64200
ν	63712.2688217289	58821.6712255282	57523.2000000000
α	3.190277719e-05	3.63707317842e-05	3.91631629710e-05
t_0	2.38559451409778	3.38412182748648	4.88499029942458
ρ	2.03259831679520	2.13938722724589	2.25279045621570
$1/\rho$	0.491981121767681	0.467423562814918	0.443893925971183
t_{j1}	21	21	28
t_{j2}	36	43	43
n	16	23	16
r	0.998063248602538	0.999267711104992	0.999476595767646
F	3603.80309527417	14322.8552686895	13363.4852428387
$F_C(1,n-2)$	17.3000000000000	14.6000000000000	17.3000000000000
$F/F_C(1,n-2)$	208.312317645906	981.017484156815	772.455794383737
S_∞	56848.1710083937	52507.9044514993	51318.7804944884
V_∞	14251.0494716063	13105.7915485007	12881.2195055116
t_{final}	67.3	64.5	62.6

Table 4. Predictions for the epidemic in Austria. Optimal values of parameters and other SIR model characteristics.

For this country the results of calculations are shown in Table 4. Fig. 2 represents the results for the prediction No. 2 corresponding to the highest value of $F/F_C(1,n-2)$. It can be seen that first cases of COVID-2019 infection were probably timely identified and sick people were isolated. This may be a reason for much lower saturation level of the epidemic in Austria (approximately 12,000 - 13,000 according to the last predictions 2 and 3). The new cases could stop to appear at the moments of time 62 -65 (see two last values in the last row of Table 4). These moments correspond to April 24-27, 2020. The average time of spreading infection $1/\rho$ could be estimated as 0.44-0.47 days and is much lower than in Italy, but higher than in South Korea, [16].

Results for Ukraine

The results of calculations are shown in Table 5. Fig. 3 represents the results for the prediction No. 6-1 corresponding to highest number of initial points which were not used for calculations. Since the epidemic outbreak occurred later than in Italy and Austria, the number of observations in Ukraine is lower and presented predictions have to be treated as very preliminary ones. It can be seen that the first case of COVID-2019 infection in Chernivtsi was probably timely identified and isolated. The epidemic outbreak was probably caused by other persons. Preliminary estimations of the saturation level are approximately 1,700 - 2,000 (according to the last predictions 2-2, 5-2 and 6-1). The new cases could stop to appear at the moments of time 57-61 (see three last values in the last row of Table 5). These moments correspond to April 19-23, 2020. The average time of spreading infection $1/\rho$ could be estimated as 0.36-0.4 days.

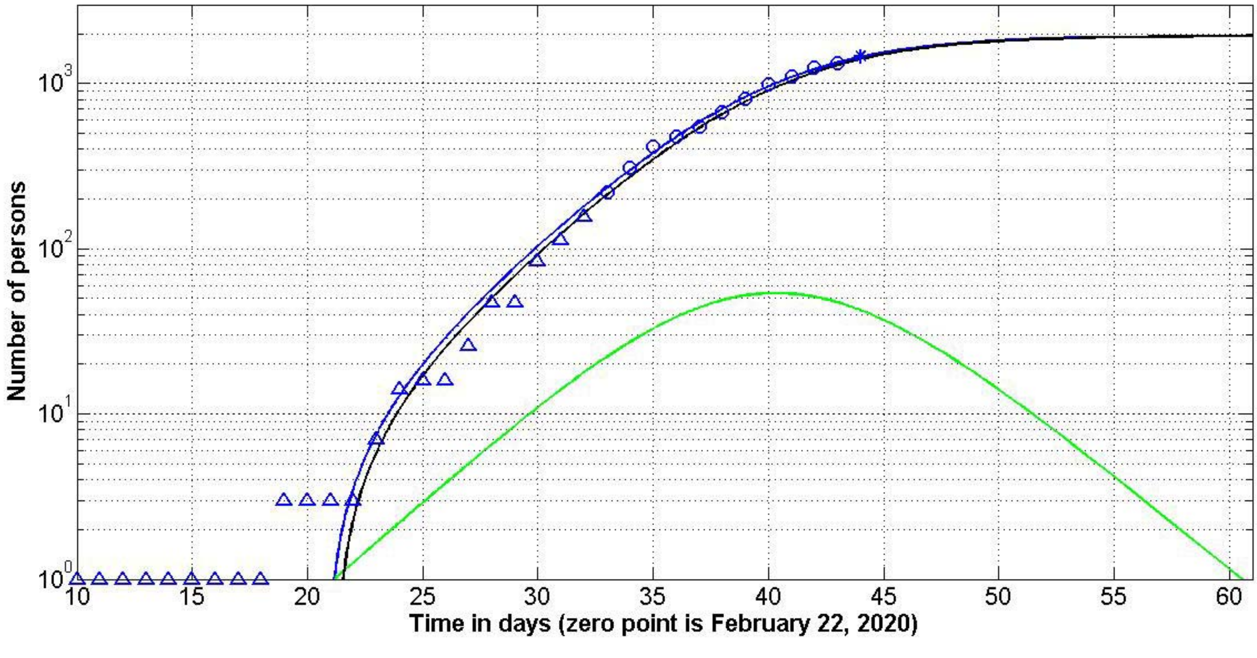

Fig. 3. Ukraine: SIR curves (lines) and accumulated number of cases (markers) versus time

Numbers of infected *I* (green), removed *R* (black) and the number of victims *V*=*I*+*R* (blue line); “circles” show the cases taken for calculations; “triangles” correspond to the cases during initial stage of the epidemic; “star” –last data point used only for a verification of the prediction.

Number of prediction	**2**	**2-1**	**2-2**	**5-2**	**6-1**
N	3692	5406.6222	7298.3283	8928.736	9835.2
ν	2768.3	4451.2941	6369.6026	8020.1295	8832.6932
α	0.0003582	0.0003336	0.0003357	0.0003422	0.0002795
t_0	20.485384	21.159201	21.902809	22.578686	21.165759
ρ	0.9916638	1.4853068	2.1382367	2.7446302	2.4690147
$1/\rho$	1.0084062	0.6732616	0.4676751	0.3643478	0.4050198
t_{j1}	25	25	25	30	33
t_{j2}	41	42	43	43	43
n	15	16	17	14	11
r	0.9984653	0.9985884	0.99876096	0.99827735	0.9972996
F	4225.6270	4948.5001	6041.85438	3474.00881	1659.6928
$F_C(1,n-2)$	17.9	17.3	16.6	18.6	22.9
$F/F_C(1,n-2)$	236.068549	286.04047	363.967131	186.774667	72.475668
S_∞	2011	3612	5518	7168	7893
V_∞	1681	1794	1781	1761	1942
t_{final}	62.1	60.5	58.7	57.4	60.5

Table 5. Predictions for the epidemic in Ukraine. Optimal values of parameters and other SIR model characteristics.

Discussion

The accuracy of any mathematical model is limited. The used SIR model is not an exception. The real processes are much more complicated. In particular, all the parameters in SIR model are supposed to be constant. If the quarantine measures and speed of isolation change or new infected persons are coming in the country, the accuracy of the prediction reduces. The accuracy of predictions increases with increasing the number of observations. On the other hand, the need for forecasts is reduced if an epidemic is stabilized.

Conclusions

The SIR (susceptible-infected-removed) model and statistical approach to the parameter are able to make some reliable estimations for the epidemic outbreaks. The accuracy of long-term predictions is limited by uncertain information, especially at the beginning of an epidemic. The long enough observations may eliminate the influence of the initial stage data and increase the accuracy of predictions. Even at limited amount of data the SIR model can be used to estimate the final size of an epidemic and its duration. The further course of COVID-19 pandemic in Ukraine, Austria and Italy will show the real accuracy of the proposed method.

Acknowledgements

I would like to express my sincere thanks to Gerhard Demelmair and Ihor Kudybyn for their help in collecting and processing data.

References

1. World Health Organization. “Coronavirus disease (COVID-2019) situation reports”. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Retrieved Mar. 14. 2020.

2. Nesteruk. I. “Coronavirus epidemic outbreak in Europe. Comparison with the dynamics in mainland China”. KPI Science News. No. 1 (2020). pp. 15-17. https://doi.org/10.20535/kpi-sn.2020.1.198043

3. Nesteruk. I. “Comparison of the coronavirus epidemic dynamics in Italy and mainland China” [Preprint.] *ResearchGate*. 2020 March. doi:10.13140/RG.2.2.19152.87049.

4. Nesteruk. I. “Comparison of the coronavirus pandemic dynamics in Europe. USA and South Korea” [Preprint.] medRxiv. 2020 Mar. doi: https://medrxiv.org/cgi/content/short/2020.03.18.20038133v1

5. Nesteruk. I. “Stabilization of the coronavirus pandemic in Italy and global prospects” [Preprint.] *ResearchGate*. 2020 March. doi: 10.13140/RG.2.2.13832.98561

6. Nesteruk. I.. Kydybyn I.. Demelmair G. “Coronavirus pandemic dynamics in March. 2020. What can we expect in April?” [Preprint.] *ResearchGate*. 2020 April. doi: 10.13140/RG.2.2.36305.28009

7. Nesteruk. I. “Comparison of the coronavirus pandemic dynamics in Ukraine and neighboring countries” [Preprint.] *ResearchGate*. 2020 Mar. doi: 10.13140/RG.2.2.31170.53448

8. Kermack. W. O. & McKendrick. A. G. “A contribution to the mathematical theory of epidemics.” *Proceedings of the Royal Society*. Ser. A. vol. 115. pp. 700–721. 1927.

9. Murray. J. D. *Mathematical biology*. 3rd ed. 2 v. New York : Springer. 2002–2003.

10. Langemann. D.. Nesteruk. I. & Prestin. J. “Comparison of mathematical models for the dynamics of the Chernivtsi children disease.” *Mathematics in computers and simulation*. vol. 123. pp. 68–79. 2016. doi:10.1016/j.matcom.2016.01.003.

11. Nesteruk. I. “Statistics based models for the dynamics of Chernivtsi children disease.” AMMODIT Conference. Kyiv. Ukraine. January 2017. *Naukovi visti NTUU KPI*. 2017. no. 5. pp. 26–34. doi:10.20535/1810-0546.2017.5.108577.

12. Nesteruk. I. "Statistics-based predictions of coronavirus epidemic spreading in mainland China." *Innovative biosystems and bioengineering*. vol. 4. no. 1. pp. 13–18. 2020. doi:10.20535/ibb.2020.4.1.195074.

13. N.R. Draper and H. Smith. *Applied Regression Analysis (3rd ed.)*. John Wiley. 1998.

14. https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf

15. I. Nesteruk. "Maximal speed of underwater locomotion". Innov Biosyst Bioeng. 2019. vol. 3. no. 3. pp. 152–167. Doi: https://doi.org/10.20535/ibb.2019.3.3.177976

16. Nesteruk. I. "Estimations of the coronavirus epidemic dynamics in South Korea with the use of SIR model" [Preprint.] *ResearchGate*. 2020 Mar. doi: 10.13140/RG.2.2.15489.40807.