Cross-sectional: BIOMEDICAL ENGINEERING

THE CORONAVIRUS DISEASE COVID-19 SPREADING PREDICTION IN UKRAINE BY MEANS OF MICROSOFT EXCEL

Yuri Palaniza[1], Halyna Shadrina[2], Mykola Khvostivskyy[3]

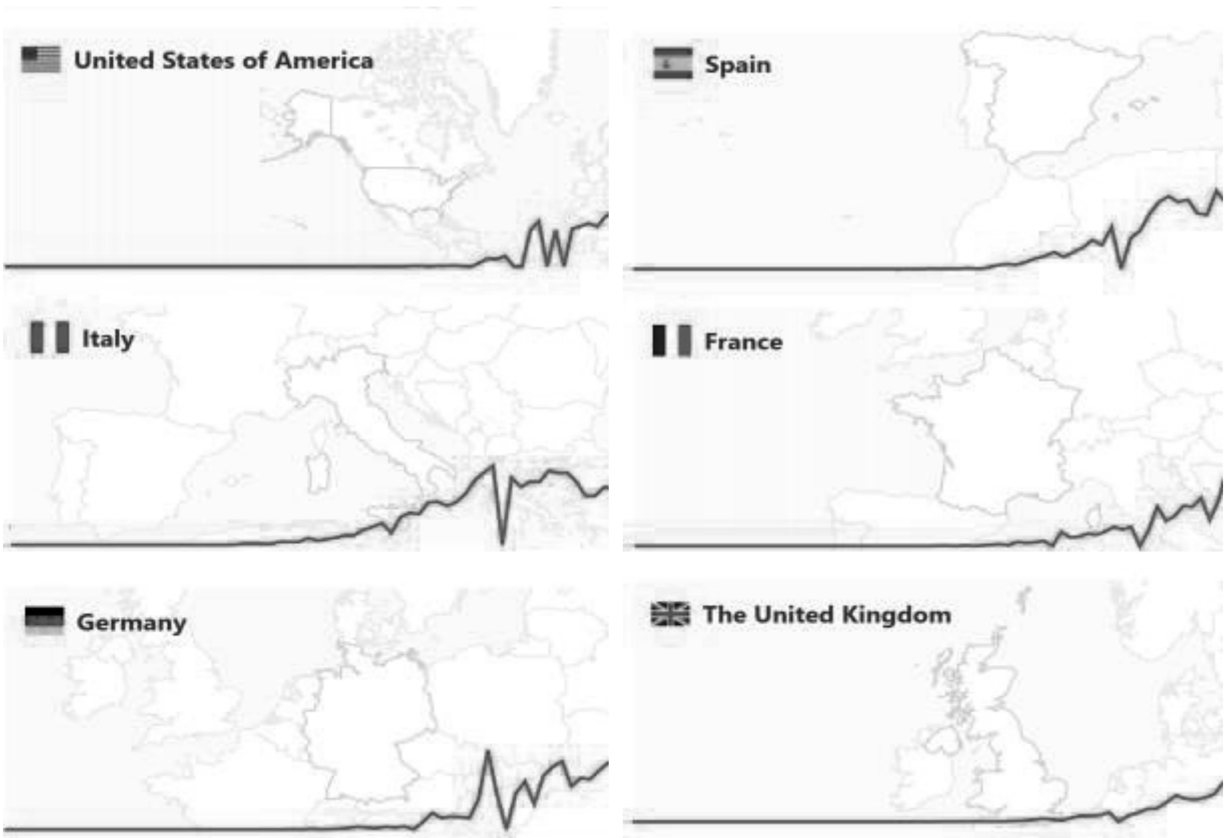
[1]*Ternopil National Ivan Puluj Technical University, Rus'ka str. 56, 46001, Ternopil, Ukraine; palaniza@ukr.net*
[2]*Ternopil National Ivan Puluj Technical University, Rus'ka str. 56, 46001, Ternopil, Ukraine; shadrinagal@gmail.com*
[3]*Ternopil National Ivan Puluj Technical University, Rus'ka str. 56, 46001, Ternopil, Ukraine; hvostivskyy@tntu.edu.ua*

Abstract: COVID-19 has completely eclipsed the morbidity, mortality, and economic impact of any other seasonal flu or communicable disease in the past century, and it is vitally important to have means that make it possible to predict its spreading. Analysis of the situation regarding the course of COVID-19 as well as diagrams of the future pandemic emphasizes their ability to change the configuration and behavior, which requires a rapid response to these changes. In view of this, an important factor is the efficiency and the need to develop a simple model, suitable for understanding even by non-specialists and with the possibility of use in the "field". Based on the Excel-project of the North-Western University, a system of automatic identification of the SIR-model (Susceptible, Infectious, Recovered) of COVID-19 was developed.

Keywords: *SARS-CoV-2, COVID-19, Microsoft Excel, Software, SIR Model, GRG Nonlinear solving method*

1. Introduction

In the context of a global pandemic, the question of the value of human life arises as never before. At the same time it is necessary to provide, if not comfortable, then at least proper conditions for the patient's stay in the hospital and provide him with medical care at the minimum allowable level in accordance with state-approved treatment protocols.



(Single space 5pt)

Fig.1 Dynamics of COVID-19 distribution in endemic countries (O. Kryvorot′ko).

At the end of the winter of 2021, Ukraine was on the rise of the second wave of SARS-CoV-2 coronavirus infection, which causes a corresponding syndrome / disease called COVID-19. Taking into account the macro- and microeconomic situation, social irresponsibility of citizens for their own safety in the environment and a number of other negative aspects, the state is extremely difficult to respond quickly and adequately to new challenges. Every effort should be made to prevent such a situation, in case of permanent development of which, medical staff will have to choose who to include in the so-called "red" category of hopeless patients, and in a total lack of resources to focus on providing priority care to other categories. patients at random. Therefore, it is critical to use all available means to smooth the peak incidence below the capacity of medical facilities. It is mathematical modeling to predict the dynamics of the epidemic can provide such tools.

2. Basic approaches to pandemic processes modeling

The development of an adequate mathematical apparatus to describe the course of pandemic processes was a primordial problem of mankind, according to the analysis conducted by O. Kryvorotko, the review of such mathematical models has the form:

D. Bernoulli (1760) - Mathematical methods for evaluating the effectiveness of smallpox vaccination methods;

J. Brownlee (1906) – Statistical approach to immune protection: the theory of epidemics (Pearson's distribution);

W.O. Kermack, A.G. McKendrick (1927) – For the first time, the "law of active masses" was applied, according to which the number of newly infected in the population is directly proportional to the product of the current number of susceptible and infected individuals. Development of deterministic SIR-model (Susceptible - Infected - Recovered):

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta IS}{N}, \\ \dfrac{dI}{dt} = -\dfrac{\beta IS}{N} - \gamma I, \\ \dfrac{dR}{dt} = \gamma I. \end{cases} \tag{1}$$

here S - is the stock of susceptible population; I - is the stock of infected; R - is the stock of removed population (both processes of death or recovery); N - is population (is considered stable).

L. Reed, U.H. Frost (1931)) - Using a number of binomial distributions to describe the number of infected individuals in each time period;

W.H. Frost (1937) - Mathematical model of the tuberculosis epidemic;

D.G. Kendall (1957) - One of the first spatial models of epidemics based on equations in partial derivatives;

H.T. Waaler (1962) - A holistic linear mathematical model (which includes 5 equations) of the spread of the tuberculosis epidemic (which includes a description of the processes of infection, the development of latent (asymptomatic) infection and disease and the subsequent spread of infection);

C.S. ReVelle (1967) - The first mathematical model of epidemic tuberculosis based on five nonlinear ordinary differential equations. The task of the economically optimal strategy of tuberculosis control was formulated;

Noble (1974), Bailey (1975), Murray, Staley & Brown (1986) – Spatial SIR model (reaction-diffusion model):

$$\begin{cases} \partial_t S = -\beta(I + \alpha\nabla^2 I)S + d_s \nabla^2 S, \\ \partial_t I = (I + \alpha\nabla^2 I)S - \gamma I + d_I \nabla^2 I, . \\ \partial_t R = \gamma I + d_R \nabla^2 R. \end{cases} \tag{2}$$

O.O. Romanyukha (2004) Development of mathematical models of tuberculosis spread, published in the territory of the former USSR - 6 equations;

J. Zhang (2005) Severe Acute Respiratory Syndrome (SARS) model in China in 2003 - 7 equations;

M. Tahir (2010) Model of the Middle East Respiratory Syndrome (MERS) epidemic in Saudi Arabia and South Korea;

T. Chen (01.2020) Mathematical model of the spread of coronavirus COVID-19 from a source of infection (bats) to humans in Wuhan Province - 14 equations;

J. Cheng (02.2020) Inverse problem for a system of 4 integro-differential equations with a delay with information about the number of infected and cured individuals in 2 weeks;

OI Krivorotko, SI Kabanikhin (02.2020) Inverse problem for a system of 6 differential equations. forecasting and identification;

R. Sameni (03.2020) Mathematical model of COVID-19 propagation with real data - 5 equations.

Existing mathematical models for the spread of COVID-19 coronavirus have a number of limitations.

• Natural births and deaths, and some other parameters that lead to changes in population.

• Differences between men and women are not taken into account (but statistics suggest that men were more vulnerable to the virus than women).

• Age ranges (but older people are known to be more vulnerable to the virus).

• Possibility of vaccination.

• Indicator of infected visitors.

• Freight and road flows.

• Indicator of social distance.

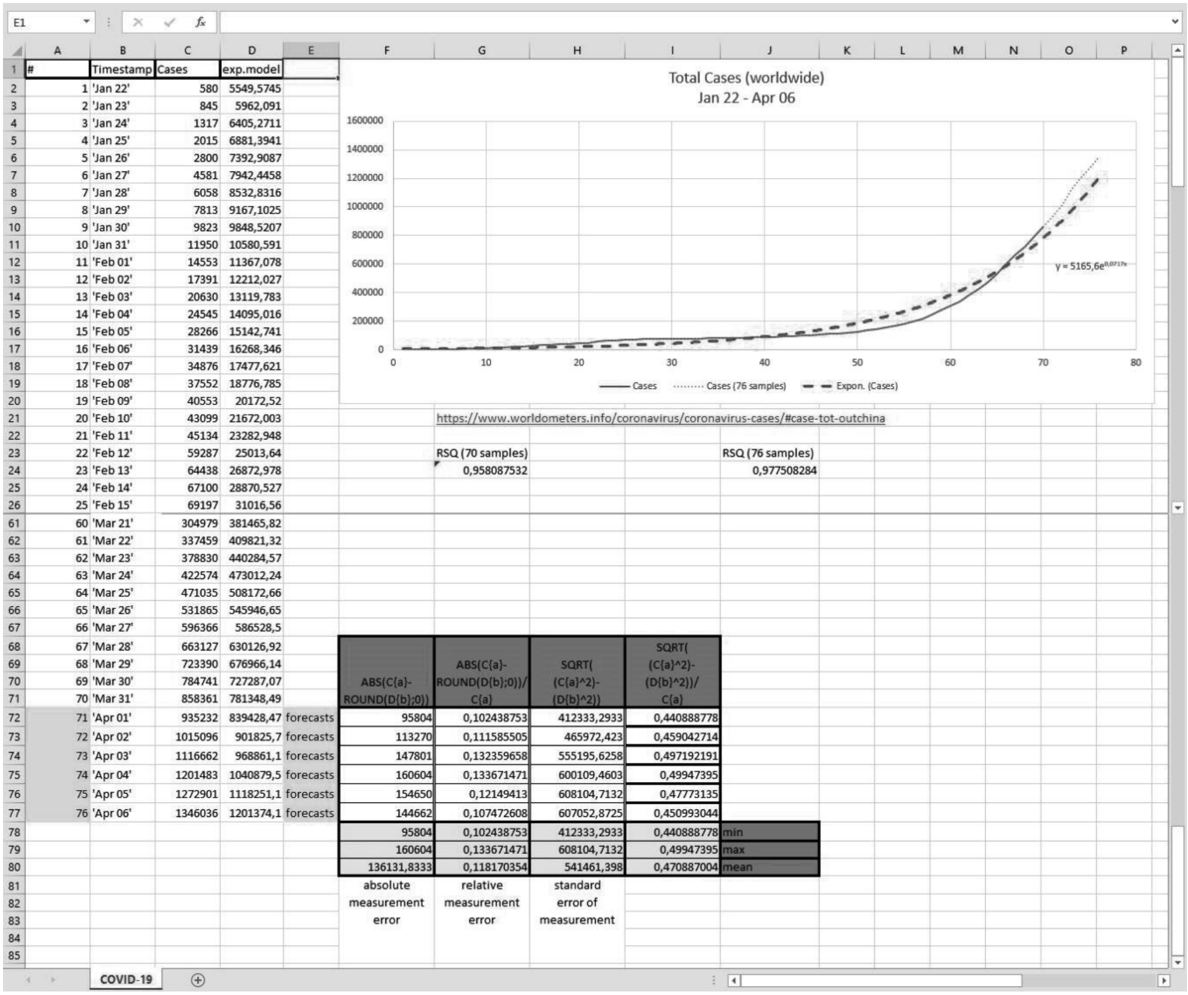
• Duration of stay in the group of infected (intensive care).

• Individual countermeasures against the spread of the virus.

3. Pandemic processes prediction

The basic approach to modeling epidemiological processes is the use of regression models. They are one of the most popular methods of predicting viral diseases. The main task of regression is to find a functional relationship between morbidity and factors that affect it, for which unknown parameters are set.

The main types of regression models in the study of viral infections are adaptive and non-adaptive. They can be used in the study of different types of diseases, to determine the information that can be used only for a particular type of model.



#	Timestamp	Cases	exp.model	
1	'Jan 22'	580	5549,5745	
2	'Jan 23'	845	5962,091	
3	'Jan 24'	1317	6405,2711	
4	'Jan 25'	2015	6881,3941	
5	'Jan 26'	2800	7392,9087	
6	'Jan 27'	4581	7942,4458	
7	'Jan 28'	6058	8532,8316	
8	'Jan 29'	7813	9167,1025	
9	'Jan 30'	9823	9848,5207	
10	'Jan 31'	11950	10580,591	
11	'Feb 01'	14553	11367,078	
12	'Feb 02'	17391	12212,027	
13	'Feb 03'	20630	13119,783	
14	'Feb 04'	24545	14095,016	
15	'Feb 05'	28266	15142,741	
16	'Feb 06'	31439	16268,346	
17	'Feb 07'	34876	17477,621	
18	'Feb 08'	37552	18776,785	
19	'Feb 09'	40553	20172,52	
20	'Feb 10'	43099	21672,003	
21	'Feb 11'	45134	23282,948	
22	'Feb 12'	59287	25013,64	
23	'Feb 13'	64438	26872,978	
24	'Feb 14'	67100	28870,527	
25	'Feb 15'	69197	31016,56	
60	'Mar 21'	304979	381465,82	
61	'Mar 22'	337459	409821,32	
62	'Mar 23'	378830	440284,57	
63	'Mar 24'	422574	473012,24	
64	'Mar 25'	471035	508172,66	
65	'Mar 26'	531865	545946,65	
66	'Mar 27'	596366	586528,5	
67	'Mar 28'	663127	630126,92	
68	'Mar 29'	723390	676966,14	
69	'Mar 30'	784741	727287,07	
70	'Mar 31'	858361	781348,49	
71	'Apr 01'	935232	839428,47	forecasts
72	'Apr 02'	1015096	901825,7	forecasts
73	'Apr 03'	1116662	968861,1	forecasts
74	'Apr 04'	1201483	1040879,5	forecasts
75	'Apr 05'	1272901	1118251,1	forecasts
76	'Apr 06'	1346036	1201374,1	forecasts

ABS(C{a}-ROUND(D{b};0))	ABS(C{a}-ROUND(D{b};0))/C{a}	SQRT((C{a}^2)-(D{b}^2))	SQRT((C{a}^2)-(D{b}^2))/C{a}	
95804	0,102438753	412333,2933	0,440888778	
113270	0,111585505	465972,423	0,459042714	
147801	0,132359658	555195,6258	0,497192191	
160604	0,133671471	600109,4603	0,49947395	
154650	0,12149413	608104,7132	0,47773135	
144662	0,107472608	607052,8725	0,450993044	
95804	0,102438753	412333,2933	0,440888778	min
160604	0,133671471	608104,7132	0,49947395	max
136131,8333	0,118170354	541461,398	0,470887004	mean
absolute measurement error	relative measurement error	standard error of measurement		

Fig.2 Prediction of the spread of the incidence of coronavirus COVID-19 in the world by means of MS Excel (as of 07.04.2020).

Adaptive regression models are mostly used during the course of the disease, to analyze and predict its future behavior. Adaptive models use the so-called concept of window width, which reflects the number of cases of manifestation of the pathogen over a period of time and are used to form an estimate of the prognosis. Depending on the width of the window, the accuracy and complexity of the model changes: as the width of the window increases, the accuracy decreases, this is due to the fact that the complexity of the function required to process a larger data stream increases. The effectiveness of this type of model is the ability to detect all local fluctuations in epidemiological indicators.

A feature of non-adaptive models is the ability to take into account previous cases of disease. Such models use all the accumulated data to construct a prognosis, as well as data from indicators of similar viruses. This type of prediction is mostly effective in using to predict seasonal virus types. A striking example of such a model is the Surfling model for detecting and predicting new outbreaks of epidemics with pronounced seasonal manifestations.

Another approach is Bayesian networks using. This type of modeling of epidemiological processes is based on the display of the model in the form of an oriented graph, on the vertices of which are variable models, and the edges correspond to the probabilistic relationships between these indicators.

The capabilities of MS Office / Excel, installed on the vast majority of personal computers running the Windows operating system, may be used for prediction the course of epidemiological processes. In this case, people without relevant skills, including epidemiologists, will be able to use the developed software..

The data of the electronic resource Worldometer [1] for the period from 22.01.2020 to 06.04.2020 were used for analysis, the last 5 of them were used for verification.

It was assumed that in the interval of explosive growth in the number of infected system can be described by the equation of the form:

$$Cases(n) = a * e^{kn}, \qquad (3)$$

where: $Cases$ – the number of cases on the n - th count, k – - speed of virus spread,, a – a constant component.

The built-in exponential regression technique was used here. Metric values R^2 throughout the interval becomes ≈ 0.98, and the relative error in the prognosis interval is smaller than 12%.

Therefore, we can conclude that the application of the technique of exponential regression, which is part of the interactive user-friendly graph data analysis interface in a spreadsheet processor environment MS Office/Excel and the corresponding mathematical model is adequate to the problem of predicting the spread of coronavirus COVID-19 in the world. This model is suitable for use by individuals without special skills in IT or Digital signal processing / System identification / regression and data fitting and without the use of specialized software on any personal computer running Microsoft Windows.

Exponential regression is a very simplified approach and is suitable for predicting the dynamics of infection spread only for a short period of time.

Derivative SIR-models operate with a large number of variables, some of which have to be set a priori, based on subjective expert assessment of epidemiological processes, which leads to low repeatability of results. Otherwise, there is a need to solve a multifactor optimization problem: the potential existence of local minima leads to "stuck" optimization process on a certain intermediate iteration and to the inadequate reality of the values of individual parameters. At the same time application of linear methods (such as Simplex algorithm) becomes impossible and it is necessary to use nonlinear (for example the Generalized Reduced Gradient algorithm). These models give acceptable results only for predicting the next wave of morbidity, while others due to sporadic processes give results that differ significantly from the true data.

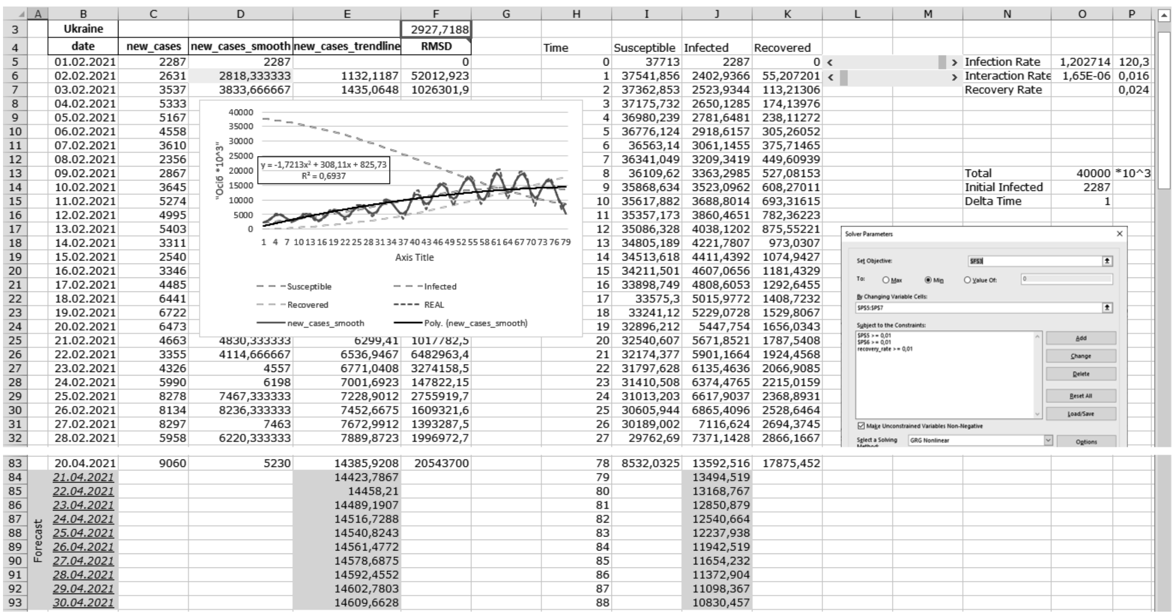


Fig.3 Prediction of the spread of the incidence of coronavirus COVID-19 in Ukraine by means of MS Excel (as of 21/04/2021).

In view of this, an important essential factor is the efficiency and the need to develop a simple model, suitable for understanding even by non-specialists and with the possibility of use in the "field". Often available tools, including Matlab, Python, R Language, are purely professional for non-medical professionals, in particular STATISTICA, Statistical Neural Network, are too difficult to master, they are paid to use, and the use of various solutions, including mobile applications, are specific. In contrast, the tabular interface is familiar to most users [2].

Based on the Excel project of Northwestern University, a system of automatic identification of the SIR (Susceptible, Infectious, Recovered) model has been developed in the interval [3] 01.02… 20.04.2021 adapted for COVID-19 in Ukraine (Fig. 3) with accuracy $\mathrm{RMSD} \approx 2927,7$ persons (Root-mean-square deviation) over the entire interval compared to real data. The automatic selection of the optimal values of the coefficients is performed by solving the optimization problem - minimization of the root mean square error between the actual and calculated values of the number of infected persons.

Figure 4 shows the algorithm for calculating the indicators [4] of infected S, currently infected I and rehabilitated R.

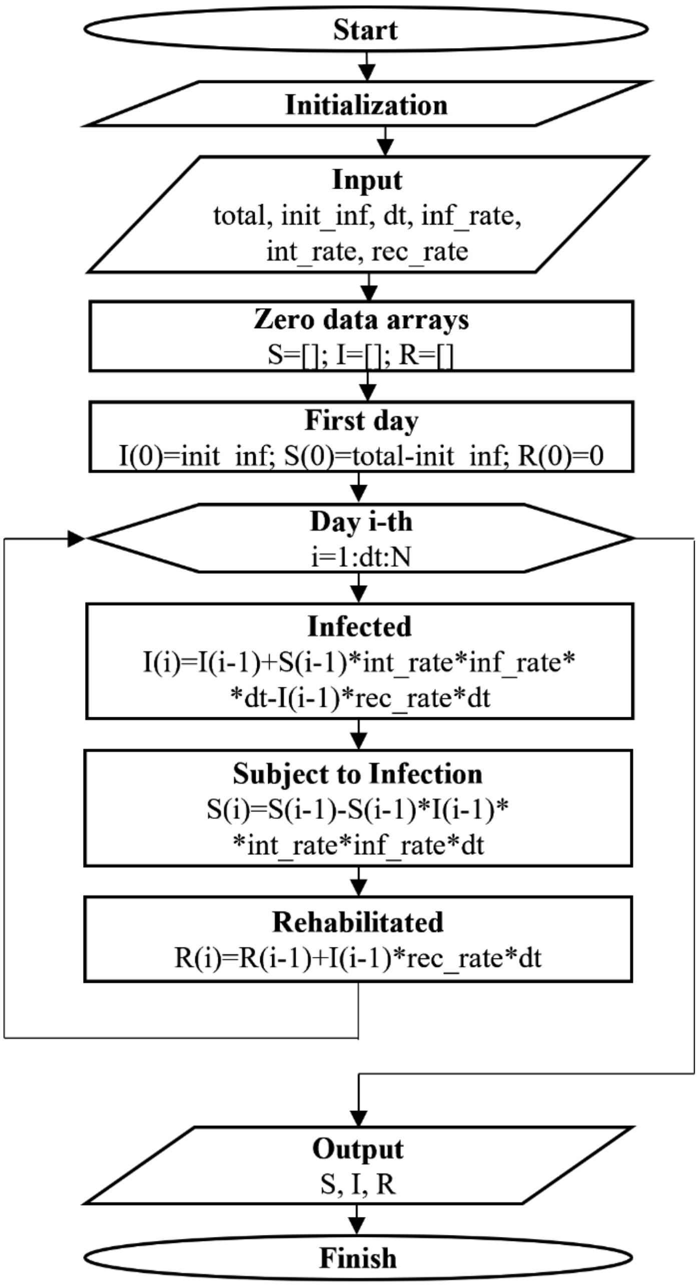


Fig.4 An algorithm for indicators calculating

In Fig.4 marked: **total** – total population; **dt** – step; **N** – number of observation days; init_inf – number of patients; **inf_rate** – infection rate; **int_rate** – intensity of individuals interaction; **rec_rate** – rehabilitation speed.

The analytical expression for the variable S (the stock of susceptible population) has the form:

$$S_i = S_{i-1} - S_{i-1} * I_{i-1} * interaction\_rate * infection\_rate * delta\_time, \quad i = \overline{0, N}, \tag{4}$$

here N - general number of days.

The analytical expression for the variable I (the stock of infected population) has the form:

$$I_i = I_{i-1} + S_{i-1} * Ii - 1 * interaction\_rate * infection\_rate * delta\_time - I_{i-1} * recovery\_rate * delta\_time, \quad i = \overline{0, N} \tag{5}$$

the analytical expression for the variable R (is the stock of removed population) has the form:

$$R_i = R_{i-1} + I_{i-1} * recovery\_rate * delta\_time, \quad i = \overline{0, N} \tag{6}$$

4. Conclusions

A system of automatic identification of the SIR model has been developed in the interval 01.02… 20.04.2021 with accuracy $RMSD \approx 2927{,}7$ persons over the entire interval compared to real data, adapted for COVID-19 in Ukraine.

Advantages of the model.

Simplicity.

The application of the smoothing technique the curve on three adjacent samples manually makes it possible to neutralize the impact of local emissions on the shape of the curve;

Using the built-in option of generating a trend line and regression equation allows both manual detranding of the signal and extrapolation / forecast of data to any number of steps forward.

Optimization of the SIR-model coefficients, as adequate to the objects of reality, by the help of automatically built-in means, allows to forecast a specific epidemic situation for any number of steps forward.

The main advantage of the approach is the convenience of an interactive user-friendly interface for data analysis in a spreadsheet environment MS Office Excel / Google Spreadsheets.

Appropriate mathematical model, adequate to the problem of predicting the spread of coronavirus COVID-19 in Ukraine, suitable for repetition without the use of specialized software by persons without special skills in the field of processing, identification of complex systems, regression analysis.

Disadvantages of the model.

A very simplified approach which allows predicting the dynamics of the infection spreading only for a short time period.

References

1. Coronavirus Cases. Total Cases (worldwide) [online cit.: 2020-04-07]. Retrieved from: https://www.worldometers.info/coronavirus/coronavirus-cases/#case-tot-outchina

2. Tatskov, O.O., Palaniza, Yu.B. (2020). The problem of predicting the spread of coronavirus COVID-19 in the world by people without special skills and without the use of specialized software on a personal computer with the operating system MICROSOFT WINDOWS. *Natural Sciences and Humanities. Topical issues*, Proceedings of the III International Student Scientific and Technical Conference. Ternopil, Ukraine.

3. Coronavirus Cases. Total Cases (Ukraine) [online cit.: 2021-04-21]. Retrieved from: https://www.worldometers.info/coronavirus/country/ukraine/

4. Suba, M. Current Mathematical Models and Numerical Simulation of SIR Model for Coronavirus Disease-2019 (COVID-19). European Journal of Molecular & Clinical Medicine 7.05 (2020): 41-54.