

# Modeling and Prediction of COVID-19 Using Hybrid Dynamic Model Based on SEIRD with ARIMA Corrections

Yaroslav Linder, Maksym Veres and Kateryna Kuzminova

Taras Shevchenko National University of Kyiv, Akademika Hlushkova Ave 4d, Kyiv, 03680, Ukraine

## Abstract

While effective prediction methods of the future dynamics of the COVID-19 pandemic can significantly improve the quality of the outbreak's containment, the number of such models specifically for Ukraine is rather low. We applied a compartment epidemiological model with heuristics along with machine learning techniques in order to create an effective method of modeling and prediction of the COVID-19 epidemic in Ukraine. The stages of the proposed method are building a SEIRD compartment model with vital dynamics, estimating its parameters, calculating and predicting the difference between the SEIRD model solution and the observed data using the ARIMA model, and adjusting model prediction using this newly obtained data on the residuals. The proposed method was tested on the data on the epidemic's dynamic in Ukraine obtained from a Ukrainian finance analytics website. The validation results indicate the method's aptitude to real-world usage.

## Keywords <sup>1</sup>

COVID-19, SEIRD, ARIMA, Hybrid Dynamic Model

## 1. Introduction

As the coronavirus pandemic continues to rattle the world, humanity craves for means to alleviate the situation if not overcome the crisis entirely. Quality estimations and predictions of future dynamics of the disease spread will ensure better prevention and thorough preparation for exacerbations of the problem (such as expected rises in infection cases after the holidays or lockdown lifts). Rational use of resources may help avoid future boiling points for the healthcare and other systems critical to the delivery of the COVID-19 response.

While the patterns of the epidemic's dynamics may be similar across countries, each country has specifics in demographics, economics, epidemic containment methods, amount of available resources, and cultural particularities, and therefore should be considered separately by researchers and scientists aiming for creating models with potential for practical usage. As shown in Figure 1, the World Health Organization reports that Ukraine has one of the highest numbers of daily increase in the number of infected individuals. Multiple models have been proposed as methods for modeling and prediction of the epidemic around the world. In contrast, the papers count for Ukraine remains relatively low. Perfecting the techniques of epidemic modeling specifically for Ukrainian statistics by independent researchers will accelerate the process of finding optimal tools and algorithms for the best possible results in models' performance. Networking and spreading awareness on novice helpful solutions and findings are crucial to this process.

The SEIR model replicates the "time-history" of any epidemic or pandemic outbreak, and it presents the model of dynamic interaction between people with four different health conditions or phases of the pandemic, namely the susceptible (S), exposed (E), infective (I), and recovered (R).

---

*IT&I-2020 Information Technology and Interactions, December 02–03, 2020, KNU Taras Shevchenko, Kyiv, Ukraine*

EMAIL: yaroslav.linder@gmail.com (A. 1); mmveres@gmail.com (A. 2); kuzminovakateryna@gmail.com (A. 3)

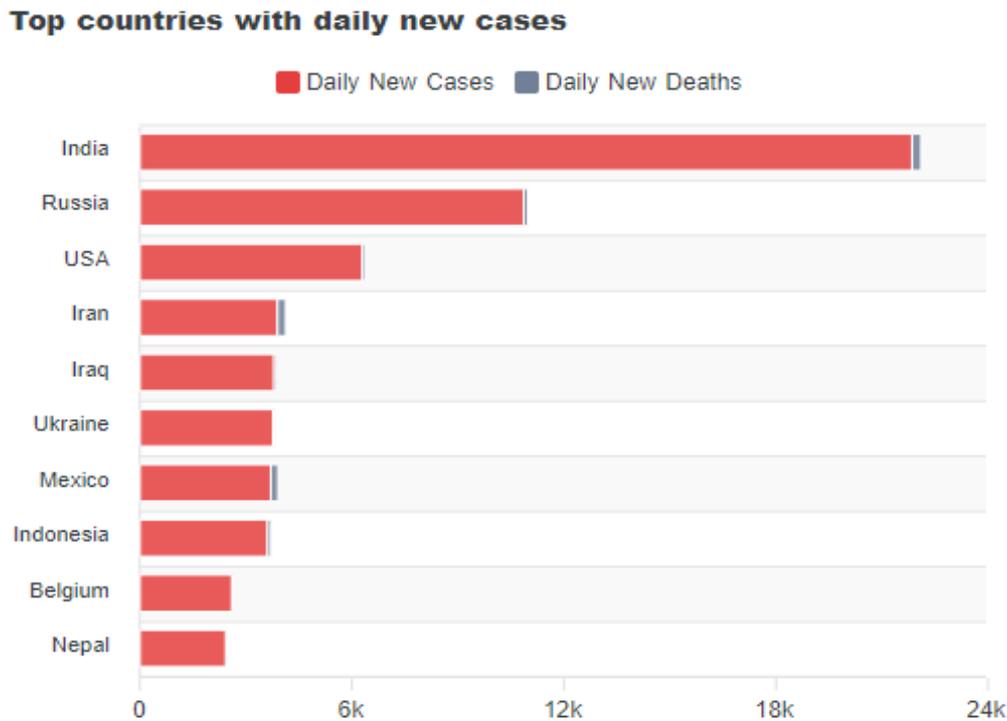
ORCID: 0000-0003-1076-9211 (A. 1); 0000-0002-8512-5560 (A. 2); 0000-0003-1236-5659 (A. 3)

© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

SEIRD model, as a generalization of the SEIR model, has an additional variable – Deceased individuals. A “Formal Characterization and Model Comparison Validation” based on the SEIRD model, which uses the data from Korea and Spain, is proposed by Casas et al. [3]. The proposed model showed the predicted parameterization with empirical evidence and a decision support system (DSS) is implemented to study the nature of the pandemic in Catalonia [3].



**Figure 1:** Rating of countries by the number of daily new infection and death cases provided by the World Health Organization

A data-driven model to predict the spread of Covid-19 for an upcoming week using the SEIRD model is studied and tested for datasets obtained from Italy, India, and Russia [2]. The proposed model [2] produces results in which the parameters are calculated from the data to plan for the future requirement of PPEs for hospital staff and healthcare devices. Contrarily, the transmission dynamics of Covid-19 were evaluated based on a SEIRD compartmental modeling approach by Mukaddes et al [4]. However, external influences such as weather, herd immunity were not considered as a part of the study. A generalized SEIR model study on the Italian Covid-19 dataset was carried out by Godio et al. [5] with parameters adjusted via Swarm Optimization Algorithm. The authors [5] claim that the method followed aims to enhance the reliability of predictions. This research is spearheading in the regions of Spain and South Korea, however, has its limitations that include the conditions of partial infections due to exposure [6], or it classifies the category of symptomatic and asymptomatic cases [7] due to the nature of the epidemic spread.

## 2. Materials and Methods

### 2.1. Database

The proposed method was tested on the data on the epidemic’s dynamic in Ukraine obtained from a Ukrainian finance analytics website [8]. The dataset includes daily information on the number of infected, recovered, and deceased individuals. The data is updated daily, enabling researchers to update model parameters frequently to achieve the highest accuracy possible. The first available observation dates back to March 3. Dataset consists of such columns:

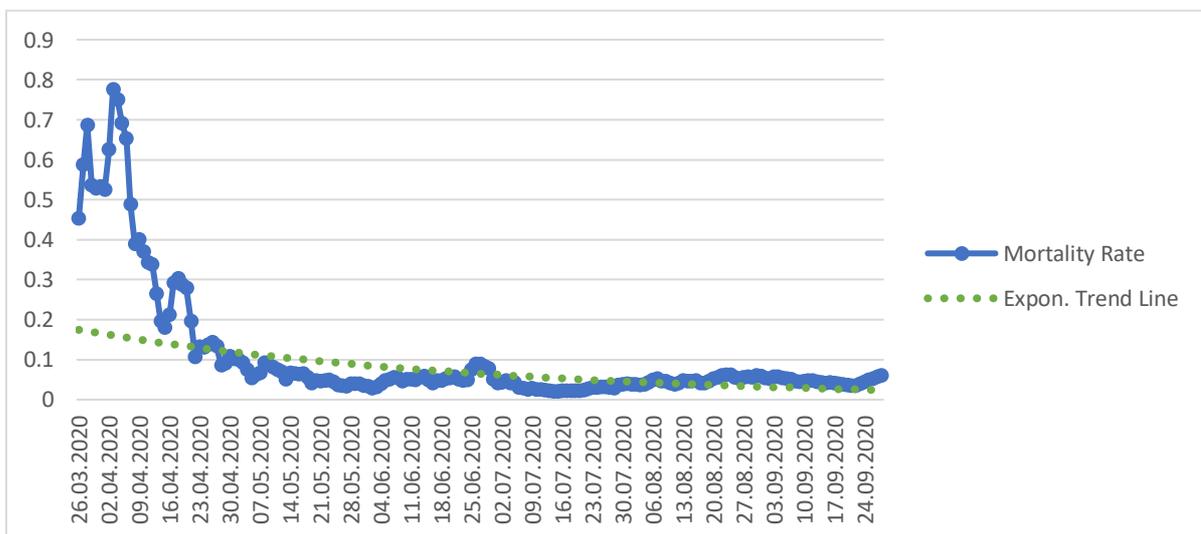
1. Cumulative infected people as of each date (the total number of diagnosed people until each date);
2. Cumulative recovered people from the start of the outbreak (the total number of no longer ill people who gained immunity until each date);
3. Deceased people from the start of the outbreak.

**Table 1**

Statistics of the COVID outbreak in Ukraine as of key dates in government responses to the COVID-19 pandemic

Date\Key statistics	Infected	Increase in Infected	Recovered	Deceased
March 12	1	0	0	0
March 23	73	10	1	3
April 6	1319	11	28	38
May 7	13691	507	2396	340
June 1	24012	664	19548	1173
July 22	60995	829	33172	1534
August 26	110085	1670	53454	2354
September 28	201305	2671	88453	3996

The key dates in the dynamics of the outbreak are stay-at-home advisories enactments and other government-enforced restrictions (March 12, March 23, April 6, July 22, August 26) and their lifts (May 7, June 1). The last day of observation used while building this method is September 28. The observed data as of those dates is reported in Table 1. The data as of later dates (up until October 19) is used for validation of the proposed method. Since lockdown and other introduced measures didn't significantly drop the outbreak's spread rate, they aren't considered in the proposed model, and basic model parameters are proposed to take as fixed. Due to the small number of cases of re-infection, all recovered individuals are assumed to have absolute immunity against COVID-19.

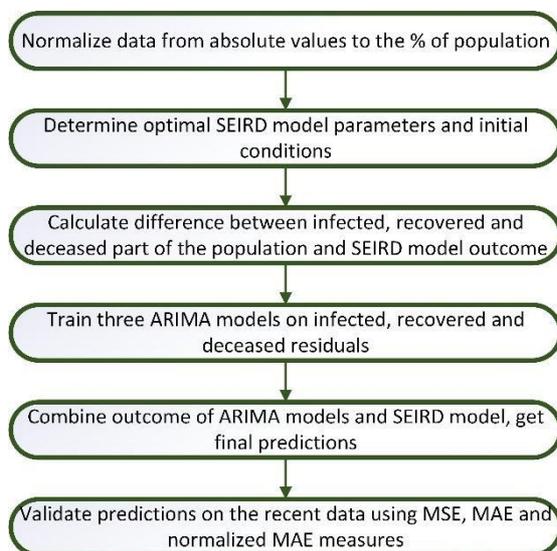


**Figure 2:** COVID-19 mortality rate dynamics and its trend line

As reflected in Figure 2, the COVID-19 mortality rate has decreased and stabilized over time, which was reflected in the dynamic model. This can be explained by continuous scientific efforts to cure the disease more efficiently as well as the proportion of asymptomatic and undiagnosed cases that aren't reflected in statistics. The relatively stable mortality rate observed in later months proves the disease to be lethal to a small portion of the population and is expected to stay at this level or slightly decrease. The data instances used while working with the model are represented in percents of the country's population.

## 2.2. The Hybrid Dynamic Model Framework

Upon investigation, we introduce a novice model based on an enhanced SEIRD model and ARIMA model. As shown in Figure 2, the stages of the proposed method are building a SEIRD compartment model with vital dynamics, estimating its parameters, calculating and predicting the difference between the SEIRD model solution and the observed data using the ARIMA model, and finally adjusting model prediction using this newly obtained data on the residuals.



**Figure 3:** The workflow of the proposed algorithm

This model consists of such stages:

1. At the first one, we estimate SEIRD model parameters using historical data, trying to lessen the difference between the model's output and observed data. This model is responsible for long-term prediction (i.e., 60 days or 100 days).
2. Calculate residuals between observed infected, recovered, and deceased percentage of the population and corresponding solutions of the SEIRD model.
3. Build three ARIMA models on the time-series of each of these residuals. Prediction of these ARIMA models will compensate residuals between the SEIRD model and historical data in order to make predictions more accurate.
4. Validate the prediction of the obtained model using the data on the number of infected, recovered, and deceased individuals as of the most recent days, data on which was not included while working with the model on previous stages.

## 2.3. SEIRD Model with Vital Dynamics and Dynamic Mortality Rate

A basic compartment model in epidemiology is the SIR model [9, 10], which studies the population's flow between three compartments: Susceptible, Infected, and Recovered. It has already been applied to the recent COVID-19 pandemic and showed good results [11]. The next level of complexity is introducing vital dynamics (birth and mortality rates) to the model [12]. Since the coronavirus disease has quite a long incubation period, it is logical to model the pandemic with another compartment – Exposed individuals who already are infected but cannot spread the virus further yet. Such model is called a SEIR compartment model. One more introduced compartment that completes our compartment structure is Deceased individuals.

A SEIRD model simulates the flow of the population between Susceptible, Exposed, Infected, Recovered, and Deceased groups (or compartments). While traditionally compartment models are built for closed systems, in this method, the total population size is not fixed due to the introduction of birth and mortality rates. This allows us to model the pandemic more accurately. The COVID-19

mortality rate is represented by an inverse exponential function with two parameters rather than a constant. Based on the analysis shown in Figure 1, it was proved to be useful to model mortality rate as an inverse exponential function, which is another heuristic to the proposed method for the same reason.

The compartments of the model are as follows:

- $S(t)$ : Susceptible individuals - stock of healthy people who may be infected; population inflow due to births is taken into account.
- $E(t)$ : Exposed individuals - virus carriers in the latent stage, during which they are not virus spreaders. Usually corresponds to an asymptomatic phase of the disease.
- $I(t)$ : Infectious individuals - virus carriers able to spread the disease to individuals in contact with them.
- $R(t)$ : Recovered individuals - stock of healthy people who are immune to COVID-19.
- $D(t)$ : Deceased individuals - population loss due to the disease, natural deaths included.

The model itself is comprised of a system of differential equations:

$$\begin{aligned}
 \frac{dS}{dt} &= \Lambda N - \mu S - \frac{\beta SI}{N} \\
 \frac{dE}{dt} &= \frac{\beta SI}{N} - (\mu + \sigma)E \\
 \frac{dI}{dt} &= \sigma E - (\zeta + \mu)I \\
 \frac{dR}{dt} &= (1 - \mu_{COVID}(t))I - \mu I \\
 \frac{dD}{dt} &= \mu_{COVID}(t)I
 \end{aligned} \tag{1}$$

with constraints at time  $t=0$   $S=S_0, E=E_0, I=I_0, R=R_0, D=D_0$  and parameters

- $\Lambda$  – population's birth rate;
- $\mu$  – population's mortality rate;
- $\beta$  – rate of virus transmission, which is the probability of transmitting disease between a susceptible and an infectious individual;
- $\sigma$  – rate of latent individuals becoming infectious (average duration of incubation is  $1/\sigma$ );
- $\zeta$  – recovery rate, which can be initially estimated as  $= 1/D$ , where  $D$  is the average duration of infection;
- $\mu_{COVID}(t)$  – death rate due to COVID-19, which is estimated by an inverse exponential formula  $\mu_{COVID}(t) = \alpha e^{-\xi t}$ .

The population size  $N(t) = S(t) + E(t) + I(t) + R(t)$  is not fixed due to its global birth and mortality rates taken into account at any given time  $t$ .

## 2.4. Parameter Estimation Using Basin-hopping Algorithm

To use the model proposed in the previous section, firstly, we need to specify its parameters so it will fit the historical data. Moreover, we estimate not only the model parameters but also initial conditions for susceptible and exposed compartments of the model. The reason of it that we still don't know the percentage of the population that is insusceptible to the virus (they will suffer from the disease in a mild form and don't infect others). Regarding the exposed population, we also don't have the exact number of exposed passengers that came to Ukraine at the time of the COVID outbreak.

As soon as the dataset consists of cumulative data, we calculated the number of currently infected individuals as a difference between cumulative infected and recovered ones. After this step, data was rescaled from the absolute numbers to the percent of the population.

To fit model parameters and initial conditions, we use the Basin-hopping algorithm [13]. This iterative heuristic algorithm is a generalization of the simulated annealing algorithm, which was inspired by molecular processes that occur in metalwork. The procedure of annealing is used to achieve the optimal molecular arrangements of metal particles. While cooling, heated material comes into shape with minimal system energy - and therefore, less or no defect. After choosing an initial

state, the algorithm picks the neighboring state and proceeds to decide on moving to it or staying and then iterates this process until finding the global optimum or reaching the iterations limit. As a generalization to simulated annealing algorithm, Basin-hopping global optimization technique randomly perturbrates coordinates and proceed to find the global optimum in a similar manner.

One of the key reasons for choosing this instrument is the algorithm's ability to reach global optima even after finding several local ones, as it is not restricted to the best candidates at each step. As a measure of quality between differential equation solution and historical data, we use MAE/mean metrics that were described and investigated in [14]. Thus, as an objective function of the Basin-hopping algorithm, we select the sum of

$$\frac{MAE(I_A, I)}{I_A} + \frac{MAE(R_A, R)}{R_A} + \frac{MAE(D_A, D)}{D_A} \quad (2)$$

where  $I_A(t)$  is the actual percentage of the population that stays infected at day  $t$ ,  $I$  is the actual percentage of the population that overcame the disease till day  $t$ ,  $D_A(t)$  is the actual percentage of the population that was deceased till day  $t$ ,  $\underline{I_A}$ ,  $\underline{R_A}$  and  $\underline{D_A}$  is the average values of infected, recovered, and deceased values over time domain,  $MAE(\cdot, \cdot)$  is calculated according to equation (2).

## 2.5. ARIMA Models for Residual Estimation

In this step, the difference between data by SEIRD algorithm and observed data is estimated and corrected using the ARIMA model (stands for Auto-Regressive Integrated Moving Average).

The structure of this model includes autoregression and moving average as the main components. The autoregression algorithm uses a certain number of past data instances (also called the number of lagged observations) to make a prediction about variable value at each new point, exploring trends and co-dependencies of observations.

Differentiation of raw data is performed to ensure stationarity of variable: each value at time  $t$  is subtracted from the value at time  $t-1$ .

The third part, moving average, also makes use of dependencies in the data, but this time between an observation and a residual error from applying the moving average algorithm to a number of lagged observations.

To each of these parts corresponds a parameter [15], where each parameter is an integer value:

p: Lag order, or number of past observations considered by the model;

d: Degree of differencing, or how many times raw observations are differenced;

q: Order of moving average, or window size for moving average algorithm.

In our case, an algorithm that finds the best set of parameters and runs statistical tests of stationarity and seasonality is used.

The obtained prediction of residuals is subtracted from data predicted by the compartment model in order to increase its performance.

## 2.6. Validation

During the validation stage, we gather new data that was not used in SEIRD model parameter estimation and ARIMA models fitting. We will use such measures of quality:

1. Mean average error, given by equation

$$MAE(y, \hat{y}) = \frac{1}{T} \sum_{t=1}^T |y(t) - \hat{y}(t)| \quad (2)$$

2. Mean squared error, given by equation

$$MSE(y, \hat{y}) = \frac{1}{T} \sum_{t=1}^T (y(t) - \hat{y}(t))^2 \quad (3)$$

3. Mean squared logarithmic error, given by equation

$$MSE(y, \hat{y}) = \frac{1}{T} \sum_{t=1}^T (\log \log (y(t) + 1) - \log \log (\hat{y}(t) + 1))^2 \quad (4)$$

4. Normalized mean average error, given by equation

$$NormMAE(y, \hat{y}) = \frac{MAE(y, \hat{y})}{\underline{(y)}} \quad (5)$$

5. Normalized mean squared error, given by equation

$$NormMAE(y, \hat{y}) = \frac{MSE(y, \hat{y})}{\underline{(y)} \cdot \underline{(\hat{y})}} \quad (6)$$

where  $\underline{(x)}$  denotes mean value of time series  $x$ . Moreover, we calculate maximum deviation between the main prediction line and two scenarios (optimistic and adverse) that are calculated from ARIMA models using a 95% confidence level. The equation of this measure is

$$MaxDev(y, \hat{y}) = \frac{|y(t) - \hat{y}(t)|}{y(t)} \quad (7)$$

### 3. Results

In this section, we will provide results of hybrid model approbation on data from the Ukrainian finance analytics website [8].

#### 3.1. SEIRD Model

In this subsection, we estimate some parameters and initial conditions of the SEIRD model using the Basin-hopping algorithm and build rough long-term predictions of pandemic development. We optimize only initial values of susceptible and exposed fraction of the population, whilst infected, recovered, and deceased initial conditions are set to zero. Global birth and death rate are also not optimized and are set according to actual values for the annual 2020 birth and death rate in Ukraine.

**Table 2**

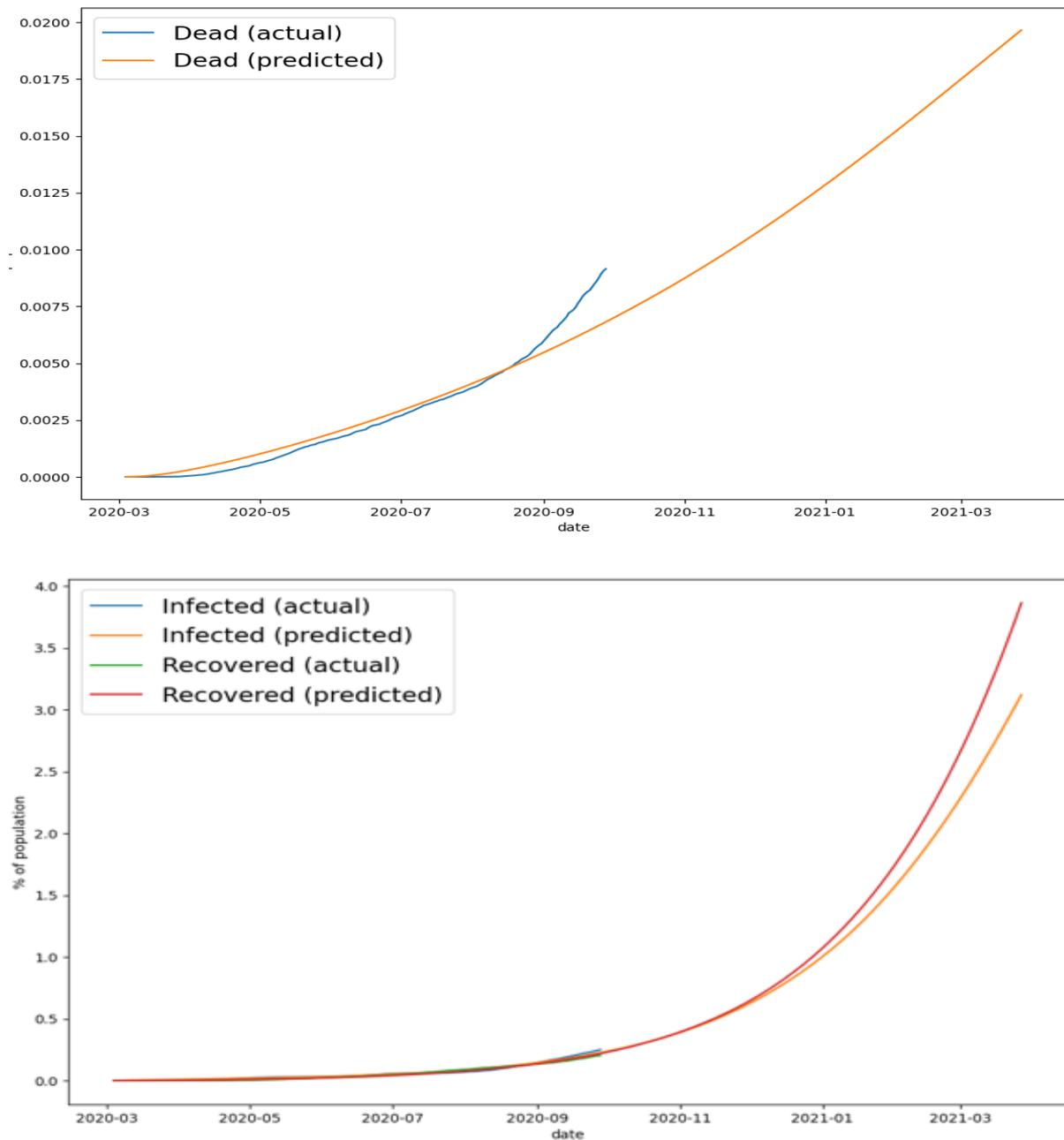
Optimized parameters and initial conditions of the SEIRD model

Parameter	Description	Minimum value	Maximum value	Optimized value
$\sigma$	Rate of latent individuals becoming infectious	0	0.1	0.0047
$\beta$	Probability of transmitting disease between a susceptible and an infectious individual	0	1	0.1529
$\zeta$	Recovery rate, which can be initially estimated as $= 1/D$ , where $D$ is the average duration of infection	0	0.1	0.0172
$\alpha$	Starting death rate from COVID	0	0.3	0.1695
$\xi$	Decaying speed of death rate due to enhancements in treatment	0	0.1	0.0121
$S_0$	Initial fraction of susceptible population	0.4	1	0.5541
$E_0$	Initial fraction of exposed population	0	0.05	0.0008

In Table 2, boundaries and optimized values for the SEIRD model parameters and initial values are shown. As we can see from the table, the initial fraction of the susceptible population is more than half of it - 55%, which correlates with recent research that most of the population will suffer from the

disease in a mild form or even asymptotically. Interestingly recovery rate is very low, which means that if a person suffers from the disease in a severe form, it takes a lot of time to recover. The rate of becoming infectious is also shallow, which proves that it takes a lot of time for the disease to be able to spread itself since acquiring a new host - the incubation period of COVID-19 is quite large. While all of the parameters have a real-life context to them and represent rates of transitions between compartments and initial conditions of the SEIRD model, they were estimated using mathematical algorithms, and that worked with available data that doesn't entirely reflect the reality. Therefore, the estimated values of some parameters such as the incubation period and recovery rate may differ from the data collected at hospitals and estimates of other researchers.

The pure SEIRD model can be used for the long-term rough predictions of the pandemic dynamic.



**Figure 4:** (a) Long-term prediction of the infected and recovered fraction of population (b) Long-term prediction of the deceased fraction of population

In Figure 4, long-term predictions for infected, recovered, and deceased fractions of the population are displayed. Based on the figures, we can conclude that number of infected people will continue to rise till summer 2021 with a relatively stable rate.

**Table 3**  
Quality measures of fitted SEIRD model

Category / measure	MAE	MSE	MSLE	Normalized MAE	Normalized MSE
Infected	$2.03 \cdot 10^{-4}$	$4.79 \cdot 10^{-8}$	$4.68 \cdot 10^{-8}$	$1.66 \cdot 10^{-2}$	$3.28 \cdot 10^{-4}$
Recovered	$8.45 \cdot 10^{-5}$	$7.94 \cdot 10^{-9}$	$7.81 \cdot 10^{-9}$	$1.04 \cdot 10^{-2}$	$1.20 \cdot 10^{-4}$
Deceased	$8.07 \cdot 10^{-6}$	$9.00 \cdot 10^{-11}$	$8.99 \cdot 10^{-11}$	$1.37 \cdot 10^{-2}$	$2.59 \cdot 10^{-4}$

Based on Figure 4 and Table 3, we can conclude that the SEIRD model fits historical data quite well. The best fit is observed for recovered and infected compartments of the model. Unnormalized measures are the lowest for the infected fraction population, which is the most informative data time-series among the studied ones.

### 3.2. ARIMA Models

At this step, we calculate residuals between the fitted SEIRD model and historical data and train ARIMA models on the residuals for each category (infected, recovered, deceased). While having its limitations [16], ARIMA can help capture any non-noisy patterns. To estimate optimal ARIMA parameters P and Q, we use the Akaike information criterion, and to estimate the optimal D parameter, we use the Augmented Dickey-Fuller test [17].

**Table 4**  
Parameters of ARIMA models for each category

Category/parameter	The order of the autoregressive model (P)	The degree of differencing (D)	the order of the moving-average model (Q)
Infected	0	2	2
Recovered	0	2	0
Deceased	0	2	0

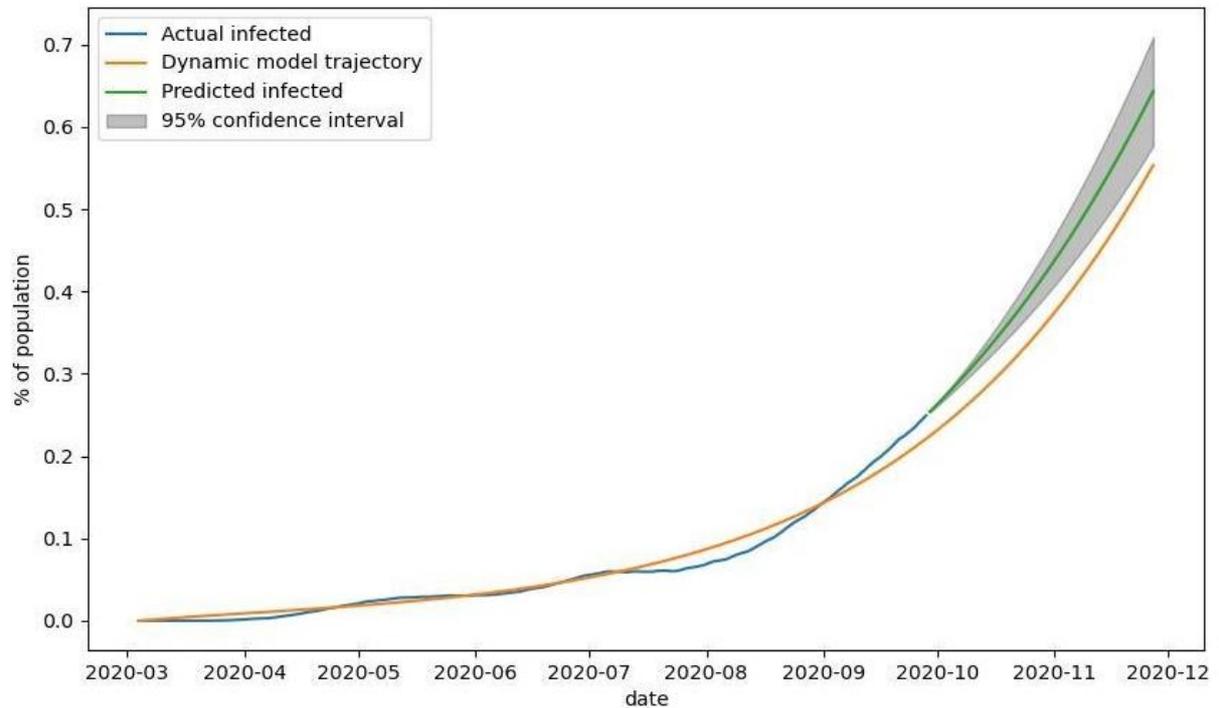
In Table 4, the estimated parameters of ARIMA models for each category are presented. Worth mentioning that for all three categories, P and Q parameters are the same, which is a good sign that tells us that the behavior of residuals time series is the same and can be simulated using similar (or even the same) models. After training ARIMA models, we evaluate predictions for all three categories 60 days ahead.

The analysis of modeling and prediction of the number of infected individuals (Figure 5) shows that the number of observed cases of the disease grew steadily during the first half of the outbreak (mid-July) and is very accurately modeled with our method.

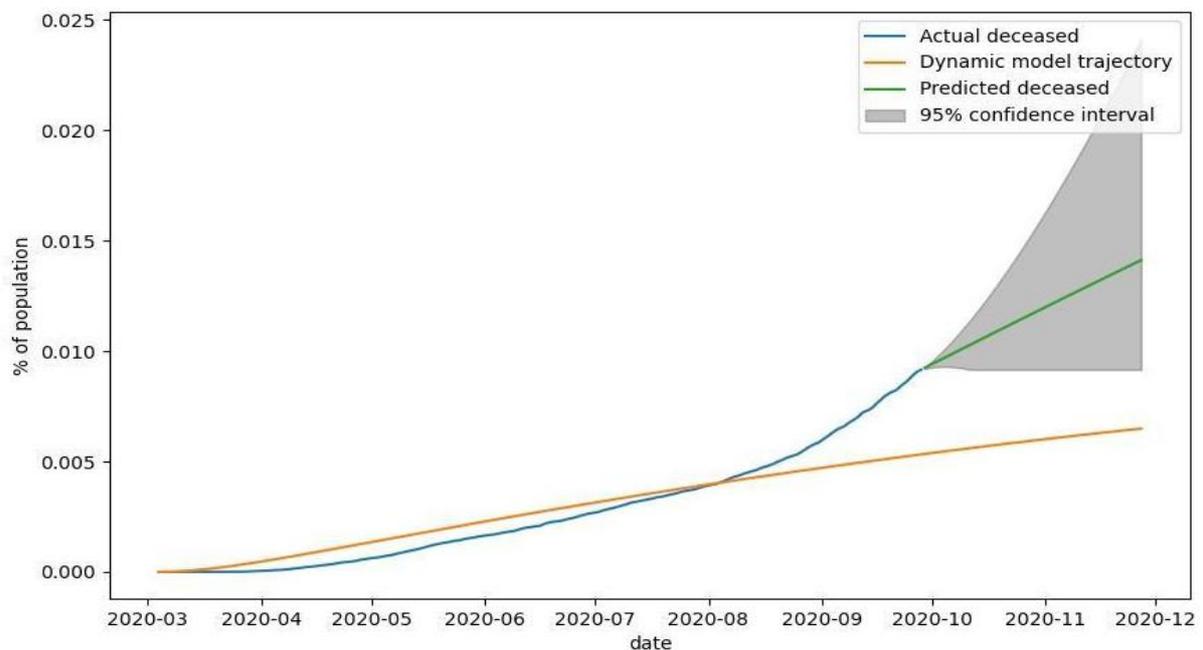
The deviation of the predicted number of infected individuals from the observed data in the second half of July and August is most likely caused by the insufficient number of tests for COVID-19 performed during this period.

The inconsistency in testing and changing levels of quarantine severity explain further deviations of observed data from the output of the SEIRD model. The prediction, corrected by ARIMA residual estimation, steadily increases, with optimistic and pessimistic scenarios (lower and upper bounds of the grey area, respectively) deviating by less than 0.1%.

As shown in Figure 6, until early August, the losses from COVID-19 are quite accurately modeled. It is safe to assume that some people who passed away due to the disease were undiagnosed or misdiagnosed.



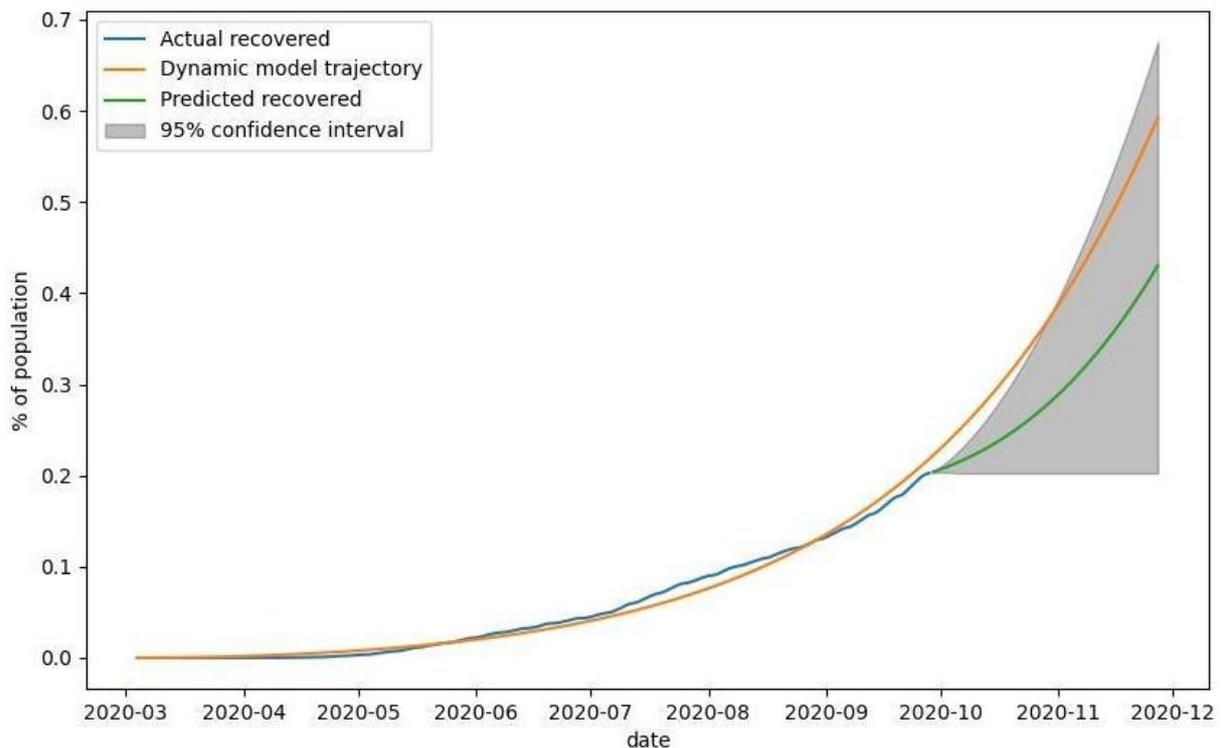
**Figure 5:** The observed number of infected individuals (blue), number of infected individuals modeled with SEIRD model (yellow), and predicted number of infected individuals (green) by SEIRD model and corrected by ARIMA residual prediction with 95% confidence interval (grey)



**Figure 6:** The observed number of deceased individuals (blue), number of deceased individuals modeled with the SEIRD model (yellow), and predicted number of deceased individuals (green) by SEIRD model and corrected by ARIMA residual prediction with 95% confidence interval (grey)

Therefore the data on those cases was not taken into account in COVID statistics, which explains the observed number of deceased people being slightly lower. In later months we observe a gradual rise – the medical system isn't well prepared for the pressure of the pandemic and struggles to cope with the growing inflow of patients.

Hopefully, there will be a decline in the COVID death rate due to the development and spreading of treatment protocols and medical research that allow selecting the most effective medicine. In the meanwhile, despite all the measures of previous months, the predicted number of deceased individuals rises quite sharply.



**Figure 7:** The observed number of recovered individuals (blue), number of recovered individuals modeled with SEIRD model (yellow), and predicted number of recovered individuals (green) by SEIRD model and corrected by ARIMA residual prediction with 95% confidence interval (grey)

The proposed method describes the observed number of recovered individuals very accurately (Figure 7) with some minor deviations, while in the future stages of the outbreak, the number of people recovered is expected to be lower than the SEIRD model suggests. It can be explained by a lack of techniques and materials to treat the patients and the already beginning congestion of the medical system of the country.

### 3.3. Validation

Validation of any method is an essential step that helps understand how the final model will perform in the future with new previously unseen data. The method was validated on the most recent data - the last three weeks (from 29.09.2020 to 19.10.2020) of the pandemic. The validation dataset was taken from the same source and therefore has the same structure.

As shown in Table 5, all measures of the prediction quality for the infected, recovered, and deceased fractions of the population are very low. Normalized MAE values show that:

1. Average difference between the actual number of infected individuals and predicted one is only 3.6%;

2. Average difference between the actual number of recovered individuals and predicted one is only 11%;
3. Average difference between the actual number of deceased individuals and predicted one is only 8.4%;
4. Based on the maximum deviation column, we can conclude that for the next 60 days starting from the last day of model training:
5. Maximum deviation between the predicted and actual number of infected individuals will not exceed 8.6% with the probability of 95%.
6. Maximum deviation between the predicted and actual number of recovered individuals will not exceed 15.4% with the probability of 95%.
7. Maximum deviation between the predicted and actual number of deceased individuals will not exceed 15.5% with the probability of 95%.

**Table 5**  
Quality measures of the fitted model for validations set

	MAE	MSE	MSLE	Normalized MAE	Normalized MSE	Max. deviation
Infected	$1.13 \cdot 10^{-4}$	$2.51 \cdot 10^{-8}$	$2.50 \cdot 10^{-8}$	$3.59 \cdot 10^{-2}$	$2.62 \cdot 10^{-3}$	8.6%
Recovered	$2.76 \cdot 10^{-4}$	$9.25 \cdot 10^{-8}$	$9.21 \cdot 10^{-8}$	$1.1 \cdot 10^{-1}$	$1.66 \cdot 10^{-2}$	15.4%
Deceased	$9.28 \cdot 10^{-6}$	$1.24 \cdot 10^{-10}$	$1.24 \cdot 10^{-10}$	$8.41 \cdot 10^{-2}$	$1.11 \cdot 10^{-2}$	15.5%

#### 4. Discussion and Conclusions

The proposed hybrid model consists of a dynamic SEIRD model with vital dynamics and decaying COVID mortality rate and three ARIMA models that cancel out dynamic model residuals and enhance prediction quality. The model was tested on Ukrainian COVID statistic data. Obtained validation results allow us to draw conclusions that the proposed hybrid model has good prediction ability and decent performance. Obtained long-term predictions reflect the general dynamic of the outbreak and are especially useful for the healthcare system workers and government officials. Obtained short-term predictions allow us not only to forecast the future number of infected, recovered, and deceased patients but only estimate forecast error under adverse or optimistic circumstances.

Key method's standouts include:

1. Using a Basin-hopping algorithm to fit parameters and initial conditions of the model for this specific disease.
2. Including into the SEIRD model exponentially decaying mortality rate, which reflects historic dynamics over the year of 2020.
3. Correction of model residuals using the ARIMA model with automatically selected parameters.

Here are some perspective ways of further development of the proposed method:

1. Parameter estimation with different algorithms and boundaries;
2. Testing the method on COVID statistics other countries;
3. Develop alternative methods for residue prediction.

Enhancing the proposed hybrid model depends on profound research results about COVID-19. That's why monitoring recent research in the field and quickly adjusting the model according to the new data is crucial.

In conclusion, the proposed method has proved its predictive capability and can be used as an effective tool for prediction and analysis of the dynamics regarding the number of infected, recovered and deceased individuals due to the COVID-19 pandemic in Ukraine. The predicted optimistic and pessimistic scenarios of the infection spread for the nearest future are very similar, so we can

conclude with sufficient confidence. Unfortunately, these conclusions give reasons to believe that the most difficult times are still ahead of us. Such results are extremely important in terms of planning disease containment measures on all levels - from governmental to personal. The analysis of obtained data indicates the forthcoming of a crisis - most importantly, in medical and economical spheres, and naturally suggests that all possible rational preemptive actions should be taken immediately.

## 5. References

- [1] Hethcote, H.W., 1989. Three basic epidemiological models. In *Applied mathematical ecology* (pp. 119-144). Springer, Berlin, Heidelberg.
- [2] Rapolu, T., Nutakki, B., Rani, T.S., and Bhavani, S.D., 2020. A Time-Dependent SEIRD Model for Forecasting the COVID-19 Transmission Dynamics. *medRxiv*.
- [3] Fonseca i Casas, P., García Carrasco, V. and Subirana, J., 2020. SEIRD COVID-19 Formal Characterization and Model Comparison Validation. *Applied Sciences*, 10(15), p.5162.
- [4] Mukaddes, A.M.M., Sannyal, M., Ali, Q. and Kuhel, M.T., 2020. Transmission Dynamics of COVID-19 in Bangladesh-A Compartmental Modeling Approach. Available at SSRN 3644855.
- [5] Godio, A., Pace, F. and Vergnano, A., 2020. SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence. *International Journal of Environmental Research and Public Health*, 17(10), p.3535.
- [6] Shi, P., Cao, S. and Feng, P., 2020. SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration. *MedRxiv*.
- [7] Shaikh, A.S., Shaikh, I.N. and Nisar, K.S., 2020. A mathematical model of covid-19 using fractional derivative: Outbreak in India with dynamics of transmission and control.
- [8] Ukrainian finance analytics website, 2013. URL: <https://index.minfin.com.ua/ua/reference/coronavirus/ukraine/>.
- [9] T. Harko, F. Lobo, and M. K. Mak, "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates," *Appl. Math. Comput.*, vol. 236, pp. 184–194, 2014, doi: 10.1016/j.amc.2014.03.030.
- [10] R. Beckley, C. Weatherspoon, M. Alexander, M. Chandler, A. Johnson, and G. S. Bhatt, "Modeling epidemics with differential equations," 2013.
- [11] W. Yang, D. Zhang, P. Liangrong, C. Zhuge, and L. Hong, "Rational evaluation of various epidemic models based on the COVID-19 data of China." 2020, doi: 10.1101/2020.03.12.20034595.
- [12] P. Shi, S. Cao, and P. Feng, "SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration," *medRxiv*, 2020, doi: 10.1101/2020.02.16.20023655.
- [13] D. Wales and J. Doye, "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms," *J. Phys. Chem. A*, vol. 101, 1998, doi: 10.1021/jp970984n.
- [14] S. Kolassa and W. Schütz, "Advantages of the MAD/mean ratio over the MAPE", *Foresight Int. J. Appl. Forecast.*, vol. 6, pp. 40–43, Jan. 2007.
- [15] A. Pankratz, "Notation and the Interpretation of ARIMA Models", <https://doi.org/10.1002/9780470316566.ch5>, 6 August 1983
- [16] S. Wang, C. Li, A. Lim, "Why Are the ARIMA and SARIMA not Sufficient", April 2019, arXiv:1904.07632.
- [17] D. Dickey, "192-30: Stationarity Issues in Time Series Models," 2005.