 doi:10.20944/preprints202206.0158.v1

STATISTICAL ERROR CALCULATION OF PANDEMIC PREDICTIVE MODELS

Muhammad Harris Nazir*
Department of Electrical Engineering
University of Engineering and Technology, Lahore
2020msee139@student.uet.edu.pk

Faheem Gohar Awan
Department of Electrical Engineering
University of Engineering and Technology, Lahore
fawan@uet.edu.pk

Umar Rashid
Department of Electrical Engineering
University of Engineering and Technology, Lahore
umar.rashid@uet.edu.pk

Muhammad Hassan Yousuf*
Department of Electrical Engineering
University of Engineering and Technology, Lahore
2020msee137@student.uet.edu.pk

1 Abstract

COVID-19 outbreak started in the Chinese city of Wuhan and spread around the globe in months due to its high contagious level. Hence, this disease posed a great challenge to researchers and mathematicians. Numerous mathematical models have been suggested to visualize the spreading speed and trends of this pandemic. In this paper, a comparison of a widely established and accepted method, the Susceptible-Exposed-Infected-Quarantined-Recovered-Death-Insusceptible (SEIQRDP) model with a newly-proposed fractional-order SEIQRDP is studied. Densely populated Countries of Asia (Pakistan, India, and Bangladesh) have been chosen as data sets and both algorithms have been applied to their data. The same comparison technique has also been used on the data of two polar countries, New Zealand and Russia, that validated our findings. Error comparison of both algorithms has been recorded in a tabulated form which shows multi-fold erroneous trends over WHO data of selected countries.

2 Introduction

The Corona Virus Disease (COVID-19) epidemic emerged from Wuhan, China, in 2019, and WHO (World Health Organization) declared it a pandemic on March 11, 2020. The disease is a respiratory syndrome and it is caused by the novel coronavirus. The outspread of COVID-19 is mostly through direct contact and the droplets of sneeze and cough. The best measures to limit the propagation of the disease are isolation, quarantine, home confinement, travel restrictions, wearing face masks, and closing people gatherings. A huge effort has been devoted by the researchers to design accurate and effective mathematical and statistical models to research the dynamic behaviors, spreading trends and development process of the disease [1; 4].

 doi:10.20944/preprints202206.0158.v1

All mathematical and statistical models, by and large, are a variant or an extension of SIR (susceptible; (at risk of getting infected by the disease); infected; recovered (insusceptible to the disease)) and SEIR (susceptible; exposed; infected; recovered) models which view the pandemic from the macroscopic perspective and describe the flow of population through three or four mutually exclusive stages of infection.
Fractional type susceptible (S), exposed (E), infected (I), and recovered (R) model incorporates the parameters like transmission rate, transition rate, and testing rates. The fractional-order COVID-19 model also shows the delay in the epidemic peak and due to the fractional-order of its model function, it flattens faster [5; 8].

SIDUR model — susceptible (S), undiagnosed infected (I), diagnosed infected (D), unidentified recovered (U), and identified removed (R) — studies the control of an epidemic [9] through testing and propose two static testing policies in order to control the epidemic:

1. Best effort strategy for testing (BEST).

2. Constant optimal strategy for testing (COST).

To predict the short-term and long-term behavior of virus relating parameters, dictionary learning is implemented to time-series data to analyze and predict the new daily reports of COVID-19 cases in multiple countries.

These studies are theoretical and do not use actual data to analyze the prevalence of specific infectious diseases. In the outbreak of COVID-19, three types of statistics are commonly available: isolated cases, recovery cases, and death cases. The conventional SEIR model cannot explain the tendency of people to be isolated and die due to the virus. Therefore, in the traditional SEIR model, the information contained in isolated cases and deaths is difficult to use.

3 Problem Statement

COVID-19 has placed a great responsibility on all of us around the world from its detection to its remediation as the globe is severely affected due to this pandemic. The most compelling strategy so far devised is its modeling through an efficient mathematical model so that tracking and monitoring of the spread of the virus would become possible and it will aid the specialized heath workers to make appropriate decisions, which are necessary to control the disease [10; 11]. A mathematical model should have the following specifications to assist us with detailed analysis and spread of the epidemic.

- Real data must be used to analyze the trend of the pandemic.
- It should consider the memory of variables and should not be local.
- It should incorporate isolated land and the impact of precautionary activities considered crucial pandemic parameters for COVID-19.
- It should provide an optimistic aid for the depiction of complicated structures notwithstanding its capability to precisely consolidate delay and memory encountered in the networks.
- It should consider time-variable parameters carefully as the countries have assumed different methodologies to handle the spread of disease as per their medical facilities and expertise.

A model consisting above mentioned specifications offers an appropriate, flexible, and reliable framework for learning the fluctuations of the pandemic.

4 Notations Used in Paper

SEIQRDP (Susceptible (S), Exposed (E), Infected (I), Quarantined (Q), Recovered (R), Death (D), Insusceptible (P)) model incorporates all the specifications for the portrayal of complex frameworks as it is constructed from seven states (sub-populations) [1].

1. Susceptible **S(t)** cases

2. Exposed **E(t)** cases, which are infected but not yet contagious
3. Infectious **I(t)** cases, which are infected but not yet quarantined
4. Quarantined **Q(t)** cases, which are confirmed, infected and quarantined
5. Recovered **R(t)** cases
6. Dead **D(t)** cases
7. Insusceptible **P(t)** cases

Each state is connected to another state with the help of these parameters:

- Unsusceptible state is connected to Susceptible state via protection rate "$\boldsymbol{\alpha}$".
- Exposed and the Susceptible States are interconnected via "$\boldsymbol{\beta}$".
- Parameter which connects Exposed and Infected states is "$\boldsymbol{\gamma}$" and it is the reciprocal of the average dormant time.
- Parameter which connects Infected and Quarantined states is "$\boldsymbol{\delta}$" which is the rate of entering quarantine.
- To connect Recovered and Quarantined states, a time dependant parameter "$\boldsymbol{\lambda(t)}$" is used, which specifies the rate of cure.

  $$\lambda(t) = \lambda_o[1 - e^{-\lambda_1 t}] \tag{1}$$

  where λ_o and λ_1 are the coefficients determined empirically.
- Death and Quarantined states are interconnected via "$\boldsymbol{\kappa(t)}$" mortality rate (time-dependant)

  $$\kappa(t) = \kappa_o e^{-\kappa_1 t} \tag{2}$$

  where κ_o and κ_1 are the coefficients determined empirically.

The parameters listed above are controlled by precautionary interventions and the efficacy of the health facilities in the region which is being investigated or analyzed. [1].

$$\begin{cases} \frac{\mathrm{d}S(t)}{\mathrm{d}t} = -\beta\frac{S(t)I(t)}{N} - \alpha S(t) \\ \frac{\mathrm{d}E(t)}{\mathrm{d}t} = \beta\frac{S(t)I(t)}{N} - \gamma E(t) \\ \frac{\mathrm{d}I(t)}{\mathrm{d}t} = -\delta I(t) + \gamma E(t) \\ \frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \delta I(t) - \kappa(t)Q(t) - \lambda(t)Q(t) \\ \frac{\mathrm{d}R(t)}{\mathrm{d}t} = Q(t)\lambda(t) \\ \frac{\mathrm{d}D(t)}{\mathrm{d}t} = Q(t)\kappa(t) \\ \frac{\mathrm{d}P(t)}{\mathrm{d}t} = S(t)\alpha \end{cases}$$

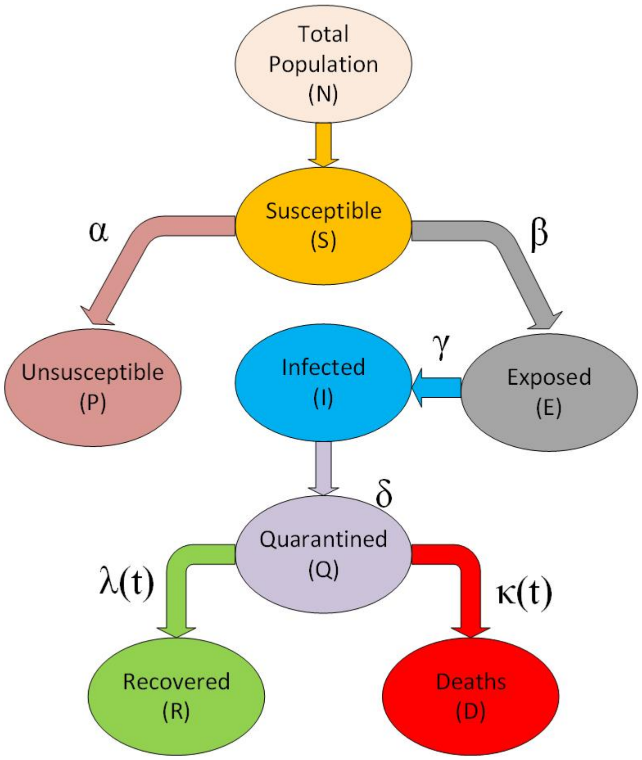
where N is the total population in the region of study.

5 Assumptions

- The natural deaths and births are not modeled means that the population of the investigated region is assumed constant.

- In pandemics, the death rate, over time, approaches zero while the rate of cure continues to increase toward a constant level.
- All other parameters are considered constant because they do not fluctuate over time.
- All infected people are supposed to be in quarantine.
- All recovered people are insusceptible to COVID-19 in the future.

Figure 1 depicts the dependency of the states on each other.



Model Layout.jpg

Figure 1: Layout illustrates the different states of SEIQRDP model based on COVID-19

6 Methods

The dynamics of all the states of a pandemic come from the mathematical model of an ordinary differential equation (ODE). SEIQRDP model uses an optimized curve fitting function in MATLAB. Ordinary Differential Equations (ODEs) are solved using Runge-Kutta Method, which is an implicit and explicit iterative method in numerical analysis. MATLAB program gets real-time COVID-19 data from Johns Hopkins Coronavirus Resource Center. Parameters of epidemic data are estimated by running a fitting optimization function and Fitting curves of real-time data are obtained.

6.1 Fractional-Order SEIQRDP Model

In recent decades, Fractional Calculus has been used extensively in the fields of biology and epidemiological modeling [12][13]. It begins with the interdisciplinary nature of the field and its effectiveness in modeling complex systems. For example, the FC method has extensively investigated the respiratory system, as well as representing biochemical characteristics, vascularity, and modeling of biological cells. [14].

Same as SEIQRDP model, the F-SEIQRDP divides the total population (N), into seven sub-population states. By the definition of fractional order derivative, it is considered that each state follows a fractional-order model [1].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0158.v1

Nonlinear FODEs in the form of matrix are considered as: $D_t^q X(t) = AX(t) + L(X)$ where,

$$D^q = \begin{bmatrix} D^{qS} \\ D^{qE} \\ D^{qI} \\ D^{qQ} \\ D^{qR} \\ D^{qD} \\ D^{qP} \end{bmatrix}$$

represents the fractional derivative operator of all seven states and "X" is the state matrix of F-SEIQRDP model

$$X = \begin{bmatrix} S \\ E \\ I \\ Q \\ R \\ D \\ P \end{bmatrix}$$

And the matrix "A" containing all the parameters of the pandemic can be formulated as:

$$A = \begin{bmatrix} -\alpha^{qS} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma^{qE} & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma^{qI} & -\delta^{qI} & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta^{qQ} & -\kappa^{qQ}(t) - \delta^{qQ}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta^{qR}(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa^{qD}(t) & 0 & 0 & 0 \\ \alpha^{qP} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

represent the parameters.

L(X) represents the non-linearity in the function of SEIQRDP and it is defined as:

$$L = S(t)I(t)[-\frac{\beta^{qS}}{N}, -\frac{\beta^{qE}}{N}, 0, 0, 0, 0, 0]^T$$

6.2 MATLAB Simulation

In this paper, the comparison of the results of the SEIQRDP and FSEIQRDP model has been made by extending their concepts to multiple countries of South Asia e.g. Pakistan, India, Bangladesh, etc. Parameter of epidemic data are estimated by the $lsqcurvfit$ function and a comparison of real-time data and fitted model is obtained. MATLAB code for country-wise simulation of SEIQRDP model and FSEIQRDP model are available on this link of google drive, and the graphical results of Active cases, Recovered cases, and No. of deaths are shown below. The layout of the implementation of the SEIQRDP model on MATLAB is shown in Figure 2. (FSEIQRDP model is implemented in the same manner by just replacing the differential equations with fractional order differential equations). Following pre-defined MATLAB functions are used in our code;

- $lsqcurvfit$ (MATLAB function for optimized curve fitting)
- $SEIQRDP$ (MATLAB function used to reenact the time-histories of a pandemic outbreak utilizing a commonly used SEIR model)
- $FSEIQRDP$(MATLAB function used to simulate the time-histories of a pandemic outbreak utilizing fractional-order SEIR model)
- $fitSEIQRDP$ (Used for the assessment of the parameters utilized in the SEIQRDP function, to show the time-variation of a pandemic wave)

6.3 Simulation Results for SEIQRDP Model

Country-wise simulation results of the SEIQRDP model showing Active cases, Recovered cases, and No. of deaths are in Figure 3, Figure 4 and Figure 5 for Pakistan, India, and Bangladesh respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0158.v1

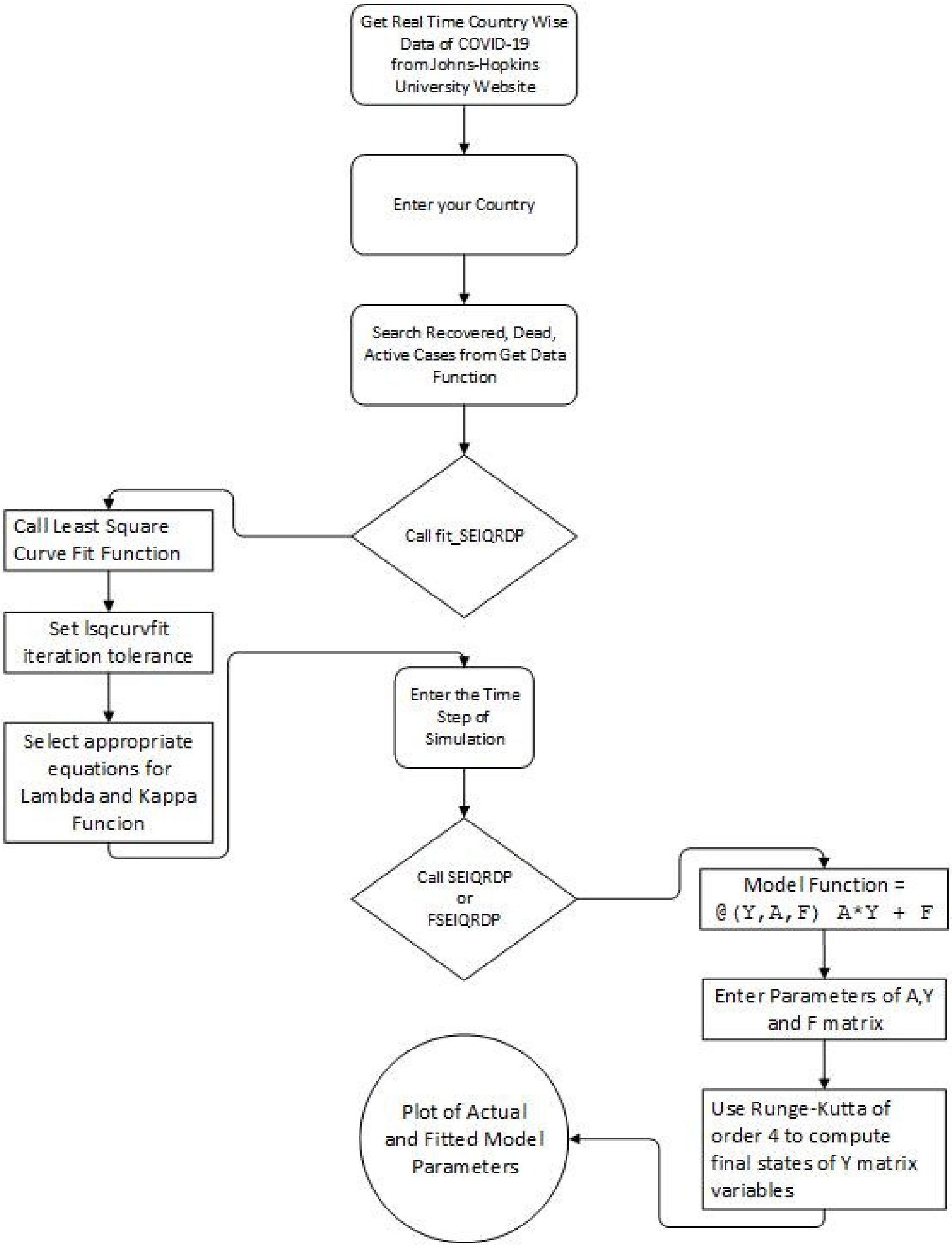


Figure 2: Coding Layout

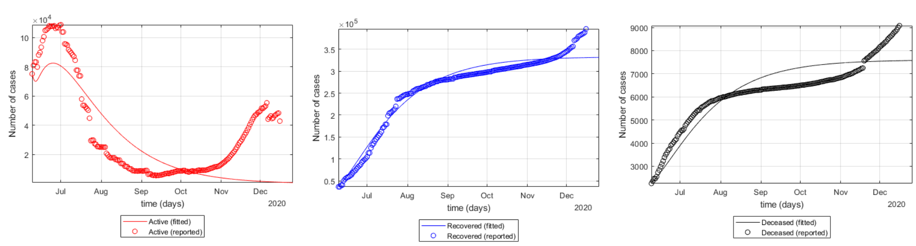
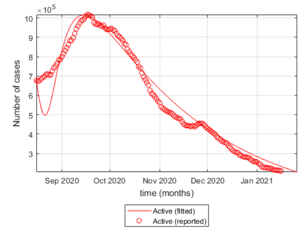
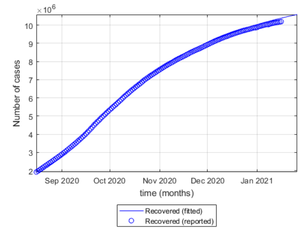


Figure 3: country = Pakistan

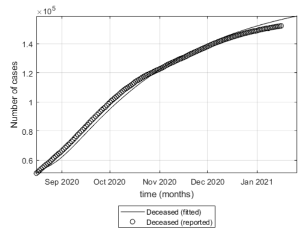
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0158.v1



Cases (India).png

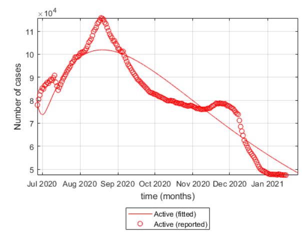


Cases (India).png

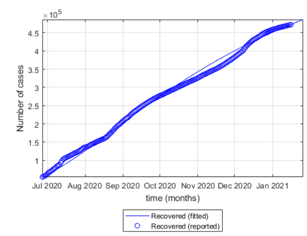


Cases (India).png

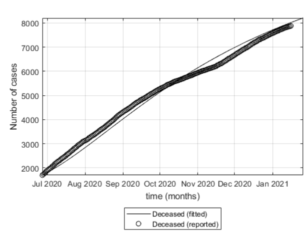
Figure 4: country = India



Cases (Bangladesh).png



Cases (Bangladesh).png



Cases (Bangladesh).png

Figure 5: country = Bangladesh

 doi:10.20944/preprints202206.0158.v1

6.4 Simulation Results for FSEIQRDP Model

Country wise simulation results of FSEIQRDP model showing Active cases, Recovered cases and No. of deaths are in Figure 6, Figure 7 and Figure 8 for Pakistan, India and Bangladesh respectively.

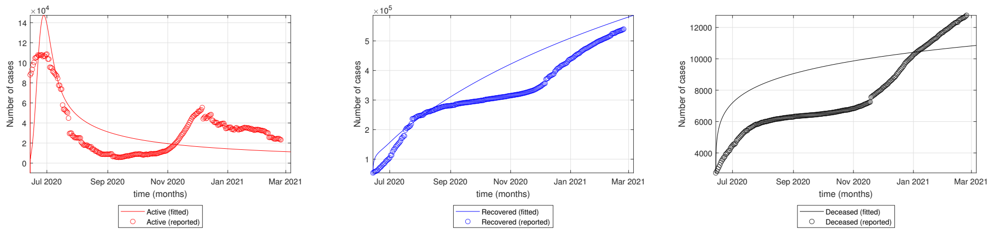


Figure 6: country = Pakistan (Fractional Order Method)

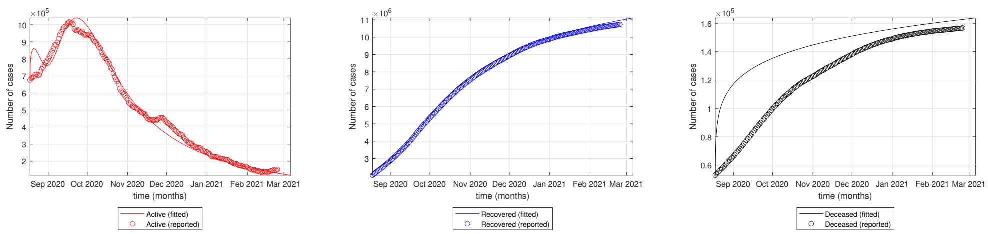


Figure 7: country = India (Fractional Order Method)

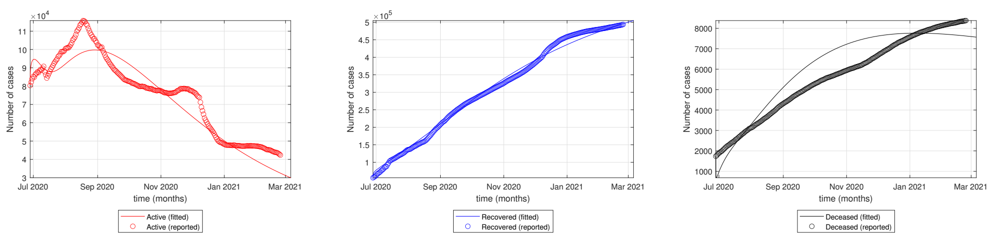


Figure 8: country = Bangladesh (Fractional Order Method)

7 Results

The evolution of the pandemic can be described with the help of mathematical models. In fact, the simulation of the pandemic elements assists us in keeping track of the escalation of virus. Also, the size of pandemics can easily be determined with the help of mathematical models and hence they may help the specialized heath workers to settle on the correct choices and limit the losses. Pandemic parameters of both models are presented in Table 7.1 and Table 7.2

7.1 Parameters of SEIQRDP Model

Country	Pakistan	India	Bangladesh
Protection Rate (α)	0.6569	0.3730	0.8751
Infection rate (β)	5	2.2625	4.999
Latent time (γ^{-1})	1	0.796	1
Quarantine time (δ^{-1})	0.06	0.0140	0.0060
λ_0	0.01	0.01	0.01
λ_1	0.0649	0.0907	0.0244
κ_0	0.001	0.001	0.001
κ_1	0.00909	0.0022	0.000371

7.2 Parameters of FSEIQRDP Model

Country	Pakistan	India	Bangladesh
Protection Rate (α)	0.6569	0.1103	0.1234
Infection rate (β)	5	0.9983	0.9724
Latent time (γ^{-1})	1	0.3295	0.1907
Quarantine time (δ^{-1})	0.0171	0.0140	0.0087
λ_0	0.01	0.07	0.01
λ_1	0.0649	0.0855	0.0244
κ_0	0.001	0.0012	0.001
κ_1	0.00909	0.00022	0.000371
qS	1	1	1.02
qE	0.98	0.95	0.98
qI	0.98	0.98	0.98
qQ	0.98	0.98	0.98
qR	0.95	0.4	0.7
qD	0.4	0.6	0.5
qP	0.5	0.8	0.399

7.3 Error Comparison of Both Models

The comparison table of both Models is given below, two types of errors have been calculated using these pre-defined formulae:

$$RMSE = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(y(i) - \hat{y}(i))^2} \tag{3}$$

and normalized error can be calculated as:

$$ReMSE = \frac{\frac{1}{l}(\sum_{i=1}^{l} |y(i) - \hat{y}(i)|)}{max(y)} \tag{4}$$

%ReMSE error can be calculated as:

$$\%ReMSE = \frac{ReMSE}{max(y)} * 100 \tag{5}$$

RMSE and ReMSE errors in active cases of both models are presented in Table 7.3:

Model	Estimation Error	Pakistan	India	Bangladesh
SEIQDP	RMSE	17555.18	48064.38	6176.459
	ReMSE	2836.69	2269.88	329.495
FSEIQDP	RMSE	19970.64	143263.0	17417.79
	ReMSE	3671.02	20166.256	2620.333

In both models, the RMSE error is almost 10 times more than the ReMSE error. Because the data set of COVID-19 is increasing every day, so the sum of the difference between actual cases and fitted cases is also increasing. It is the main cause of huge numeric values in the error table. One way to analyze these results is to calculate the percentage error of both cases.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0158.v1

8 Simulation Results of New Zealand and Russia

Both of the above mentioned models have been applied on two polar countries countries and their fitted curves and error comparison are given below:

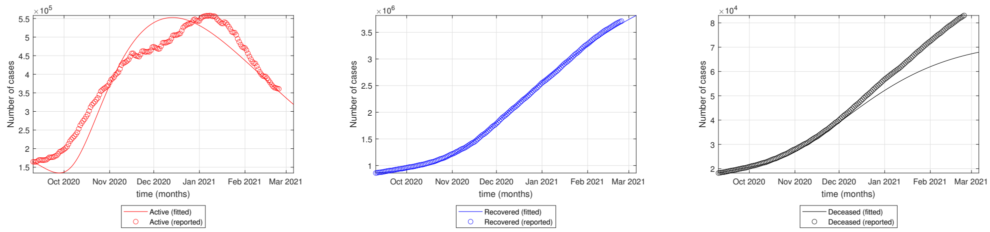


Figure 9: country = Russia

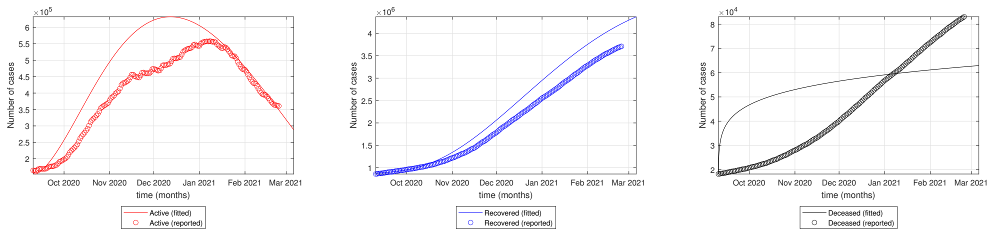


Figure 10: country = Russia (Fractional Order Method)

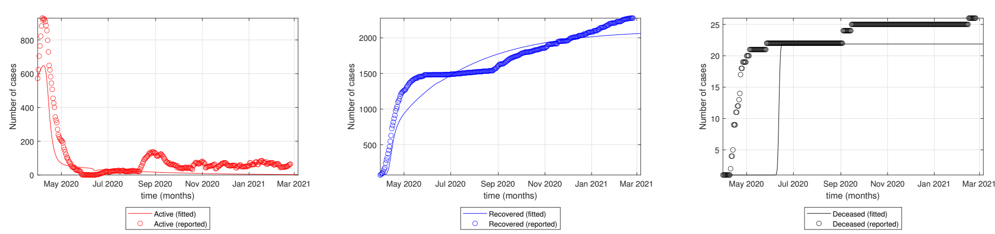


Figure 11: country = New Zealand

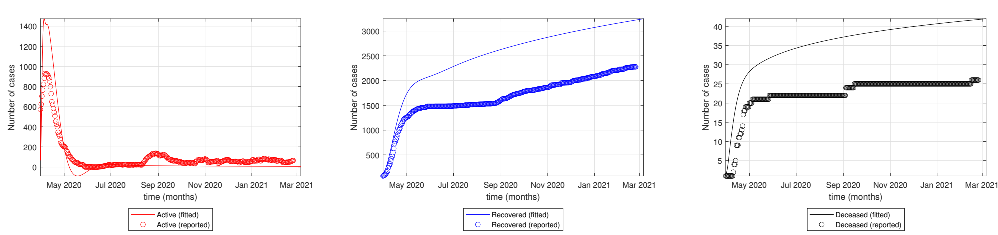


Figure 12: country = New Zealand (Fractional Order Method)

8.1 Error Comparison between New Zealand and Russia

Errors in active cases of both models are presented in Table 8.1:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0158.v1

Model	Estimation Error	New Zealand	Russia
SEIQDP	RMSE	88.5747	35634.88
	ReMSE	8.4451	2275.108
FSEIQDP	RMSE	131.8153	206714.54
	ReMSE	18.7032	76558.50

9 Results of % ReMES

The results of %ReMSE of active cases are presented in Table 9

Model	Pakistan	India	Bangladesh	New Zealand	Russia
SEIQRDP	2.611	0.223	0.2846	0.9090	0.4076
F-SEIQRDP	3.379	1.9814	2.263	2.0133	13.71

From the table above, it is evident that F-SEIQRDP has more % ReMSE Error as compared to the conventional SEIQRDP model. In some countries, the fractional-order method does not converge to an optimal solution which leads to poorly fitted curves, for example, while analyzing the data of Russia, multiple combinations of initial guesses are tried, but the F-SEIQRDP algorithm neither converges nor gives us the best-fitted curves.

10 Conclusion

From the results presented in this paper, it is clear that Fractional Order SEIQRDP is not a better algorithm to study the behavior of COVID-19. But to get more accurate results from this algorithm, the following points should be kept in mind:

- Most of the time the Fractional Order SEIQRDP algorithm does not converge to a local minimum point.
- If our initial guess matrix of fractional parameters is close to the actual values, then the algorithm converges fastly and provides us better-fitted curves, but still not better than SEIQRDP Model.
- In some cases, Fractional Order Method gives us poor fitted curves (very large error) because our initial guesses are not close to the actual values. So this leads us to the draw-back of the Fractional Order Method that the accuracy of the algorithm is immensely dependent on the accuracy of our initial guesses of parameters.
- In the case of very large or very small data sets Fractional Order Method is not as effective as the conventional SEIQRDP model.
- We should carefully consider the parameters of time because medical facilities of countries, the expertise of the front line workers, the ability to perform COVID-19 tests per day, and smart strategies to mitigate the effects of pandemics (e.g. Lock-down, Quarantine, etc.) are different for each country.
- In Fractional-Order SEIQRDP Model computational time is more than the standard SEIQRDP model.
- As the COVID-19 data-set is increasing day by day, even if our initial guesses are close to the actual values, the fractional-order method finds it difficult to converge to a local minimum. Hence it is only applicable to medium data sets.
- In case of multiple waves of the pandemic, the Fractional Order Method provides us poor fitted curves than the normal SEIQRDP model.

11 Future Directions

- A comparison of the two models has been done in this paper. In the future, these pandemic curve-fitting studies can be used to predict the upcoming peak of the pandemic. This can be achieved with the help of extrapolation algorithms.

- As suggested in this paper, the accuracy of the fractional-order method is dependent on its initial guesses, in the future an appropriate and result-oriented method can be suggested to calculate the educated guesses for the fractional-order method.

References

[1] Bahloul, Mohamed A., Abderrazak Chahid, and Taous-Meriem Laleg-Kirati. "Fractional-order SEIQRDP model for simulating the dynamics of COVID-19 epidemic." IEEE Open Journal of Engineering in Medicine and Biology 1 (2020): 249-256.

[2] Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J. Med. Virol. 2020;92:418–423. doi: 10.1002/jmv.25681.

[3] Spaan WJM, Cavanagh D, Horzinek MC. Coronaviruses: structure and genome expression. J. General Virol. 1988;69:2939–52. doi: 10.1099/0022-1317-69-12-2939.

[4] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in china, 2019. New Engl. J. Med. 2020;382:727–733. doi: 10.1056/NEJMoa2001017.

[5] He, Shaobo, Yuexi Peng, and Kehui Sun. "SEIR modeling of the COVID-19 and its dynamics." Nonlinear Dynamics 101.3 (2020): 1667-1680.

[6] Kock RA, Karesh W, Veas F, Velavan TP, Simons D, Mboera LEG, Dar OA, Arruda LB, Zumla A. 2019-ncov in context: lessons learned? Lancet. Planet. Health. 2020;4:87–88. doi: 10.1016/S2542-5196(20)30035-8.

[7] Ceraolo C, Giorgi FM. Genomic variance of the 2019-ncov coronavirus. J. Med. Virol. 2020;92:522–528. doi: 10.1002/jmv.25700.

[8] Fanelli D, Piazza F. Analysis and forecast of covid-19 spreading in china, italy and france. Chaos Solitons Fractals. 2020;134:109761–109761. doi: 10.1016/j.chaos.2020.109761.

[9] Owusu-Mensah, Isaac, et al. "A fractional-order approach to modeling and simulations of the novel COVID-19." Advances in difference equations 2020.1 (2020): 1-21.

[10] Niazi, Muhammad Umar B., et al. "Modeling and Control of COVID-19 Epidemic through Testing Policies." arXiv preprint arXiv:2010.15438 (2020).

[11] Huang, Chaolin, et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." The lancet 395.10223 (2020): 497-506.

[12] Ahmed, E., and A. S. Elgazzar. "On fractional order differential equations model for nonlocal epidemics." Physica A: Statistical Mechanics and its Applications 379.2 (2007): 607-614.

[13] Area, Ivan, et al. "On a fractional-order Ebola epidemic model." Advances in Difference Equations 2015.1 (2015): 1-12.

[14] Hilfer, Rudolf. "Fractional time evolution." Applications of fractional calculus in physics (2000): 87-130.