

Coronavirus Epidemic Spreading Prediction combining random walk with susceptible–exposed–infected–recovered–deceased (SEIRD) model

Zhengxi Wang^{1,*}, Yujian Yuan²

¹ School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215000, China

² Department of Mathematics, Boston University, Boston, United States

* Corresponding Author Email: Zhengxi.Wang20@student.xjtlu.edu.cn

Abstract. Since the breakout of COVID-19 pandemic in December 2019, effective modelling of the spreading of the virus has become an essential reference for the epidemic controlling. In a bid to solve the problem of Epidemic prediction, susceptible-exposed-infected-recovered (SEIR) model are widely applied. However, this model seems lack the ability to handle random events which may occur during the spreading of the pandemic and the ability to simulate the pandemic spreading between different subdivided regions. Therefore, we propose an early version of susceptible–exposed–infected–recovered–deceased (SEIRD) model that combines the classic compartmental concepts of SEIRD and the random walk methodology to forecast COVID-19 in real time. Specifically, this study will focus on improvement of the exposed–infected part of SEIRD model. First, the exposed–infected section of SEIRD model will be applied to each subdivided regions separated. Then, instead of entering infected–recovered part directly, the infected of each district will be selected and sent to linked districts by random walk system to mimic the commuting and irregular trips between regions. Eventually, after the re-distribution of infected patients, the model will enter the infected–recovered section. This argued model adopt the SEIRD model to forecasting of virus spreading between small regions and taking irregular moving of citizens into consideration via random walk system, thus provide an effective reference for countries which aim to respond to the post-epidemic era.

Keywords: COVID-19, Spread of pandemic, Random walk, SEIRD model.

1. Introduction

1.1. Background Information

Emerged in early 2020, COVID-19 has caused a worldwide pandemic over the past three years, which infect over eight million people in end of 2020 and one hundred million in the start of 2021 [1][2]. Confronted with such a gravity epidemic situation, it is very essential for the limited public healthy capacity and prevent recourses to be planed and allocated scientifically with an intention of arresting the spread of the virus. Therefore, an effective model which could forecast the trend of the spread of the disease in the early future becomes a necessary reference of the successful preventive activities, for instance the planning hospital resources. Typically, a well-known model is always applied in the epidemic estimation, which is susceptible–exposed–infected–recovered–deceased (SEIRD) model. However, the original model suffers from a few limitations caused by the parameter transmission rate β when the rate is a spatial-temporal irrelevant constant variable. Therefore, several studies have been published to handle this problem. Thus, before proposing our model which combines statistical approaches with SEIRD, we will first review the gist of these reports.

1.2. Literature Review

Various improvements have been made based on the SEIRD model. Some research tries to improve the regional adaptability of this model. One report from the US which collects data from different regions including US, California alone and Japan, proposed that on the epidemiologic parameters may differ owing to the spatial difference, for instance the basic reproduction number of US at last doubles compared with that of Japan [3]. This indicate that while transmission rate β may

be constant variable, it still can be assigned with different tail-made value according to the different regions so that the forecasting can be more accurate. Another 2021 paper found an alternative solution which successfully turn epidemiologic parameter into a dynamic variable and take spatial-temporal inhomogeneities (e.g., rivers and highways) into consideration [4]. Specifically, the research first captures the dynamics of pandemic, including local nature and population density, and then apply them to the SEIRD model. As a result, the simulation using above method show a high agreement with the reported data in Lombardy.

Apart from the regional adaptability, some studies manage to turn the temporal -invariant variables into a time dependence variables as well. Leded by Istanbul Technical University of turkey, a team integrates the SEIRD model with "Deep Assessment Methodology (DAM)" which is developed from fractional differential equations [2]. Via this approach, the prediction in a specific time can be simulated by a previous value, taking time factors into consideration as well. Another essay also notices the time-invariant limitation of SIR model. Therefore, this study builds the model based on time series via Autoregressive Integrated Moving Average Model (ARIMA).

A paper from China proposed an improved prediction model based the SEIRD by take vaccinees into consideration [5]. This implies apart from the spatial-temporal factors, other elements such as public health power and policy can affect the transmission rate β as well, when the ability of unvaccinated and vaccinated infected patient to infect the susceptible varier.

1.3. Thesis Statement

The literature reveals that while the original SEIRD model is limited by the spatial-temporal irrelevant constant variables, improve which enabling a more accurate prediction still could be made when we combine the initial model with other mathematical approaches. Meaning while, similar conclusion also has been made by another two papers that a short-term forecast with four week or 30-Days is adequate for the purpose of arresting spreading of virus. Nonetheless, there still exists some pointer that these above reports may miss [6] [2]. First, their improve SEIRD models could only be applied in relatively large geographical divisions level, which indicates it can be hard to get a specific prediction of a sub-division within a large geographical divisions level, for example the county within the states. Other problem is that while multiple methodical methods have been used to get a more precise parameter according to the time and place, we still cannot neglect the impact of random events, which means that the actual value of these parameters should be fluctuate around the forecasted value of parameters instead of the actual value is equal to the forecasted value.

Therefore, we propose to introduce gossan random walker as a stochastic approach to a SEIRD model. The gossan random walker will have two roles in our short-term simulation. First, the random walk will manage to simulate the fluctuation of actual values around the forecasted values of parameter to mimic the effect of random event. Another role is that the random walker will mimic the numbers of occurrence of irregular moving of citizen between linked subregions. By the combination of two method, our model not only could simulate the pandemic data within a single subregion, but also simulate the population moving (numbers of occurrence of irregular and regular moving of citizens) between linked subregions within a large region.

The general framework of this study will be presented as above. First, related works will be reviewed to find the limitations of SEIRD model that have been overcome and remain unsolved. Secondly, detailed explanations of our model and innovations will be putted in the methodology section. Then, we will prove the practicality of the proposed approach in the experiment part. Eventually, discussion and conclusion section will review the gist of our whole paper and point out the direction of possible further research.

2. Related Work

In this section, we will first explore the previous studies on random walks so that to have a better understanding of what random walk can do in different applications. Then diverse improvement suggested by various publication will be evaluated and compared with our model.

2.1. Random walk Literature

Random walk which is a random process consisting of random steps in the math space (whether on lattice or not), has been widely applied to simulation of object randomness, since it was introduced in 1905. A comprehensive investigation which is done by a 2020 report, shows that random walk is the core of Collaborative Filtering as well as recommender system where the potential preference of customers will be predicted via the measuring the similarity of random walking on the association graph[7]. Other team headed by Yan suggests that under the effect of COVID-19 Pandemic, the random walk model could fit the stock index the in a short-term where random walk would be used as a probabilistic method to decide whether an individual stock may go up or down [8]. In brief, as is unillustrated by both papers, random walk is generally be applied on the problems regarding probabilistic problems. Therefore, it is safe to deduce that this model can be used to simulate the fluctuation of actual values around the forecasted values of parameters of SEIRD, as the fluctuation of values also can be viewed as the various occurrences of actual values around the expected ones with different probabilities.

2.2. SEIRD model Literature

As a model been the most applied for the COVID – 19, while the original SEIRD model is consisting of constant parameters which is independent from the effects caused by the spatial-temporal difference and other factor (e.g., public healthy power difference), several improvements have been made to this model to push it adopt to the dynamics of pandemic. Most study focused on how to turn the static SEIRD model into a dynamically one. One 2020 report suggest changing these constant parameters into time dependence functions [2]. Another research focusing on Lombardy region not only introduce partial-differential-equation to collect temporal variables but also take spatial variables which represent local nature inhomogeneity into account, which deduce the parameters of SEIRD parameters could be represented by a spatial-time function as well [4]. Aside from the papers exploring the time and space variables, there also exists study focused on the impact on the SEIRD caused by human activities and other public factors. Head by a Colombia researcher, a team introduce machine learning to analyse the parameters and manage to find the relation between these SEIRD variables and social factors, including poverty index, number of people over 65, morbidity rates, average age [9]. Then, another report mainly investigates the explore the vaccine influence on the SEIRD model, where they find the transmission rate β of the vaccinated and unvaccinated varies [5]. In conclusion, all these mentioned studies focused on how to improve the prediction accuracy of SEILD model by applying different mathematical models or by taking some special factors into consideration.

However, instead of forecasting the spreading of pandemic, a distinctive report adopts SEIR model (a simplified version of SEIRD model) to evaluation the possible economical cost when taking different quarantine strategy [10]. Another study also regards regional lock-down and maintaining social-distancing play the major role in arresting pandemic [11]. Nonetheless, these effective administrative measures usually achieved with the cost of public fiscal expenditure as well as interrupting current economic functioning. This implies when these measures are applied in large region, the economic cost would grow as well. Therefore, it is more rational to just implement these administrative measures within small regions which is more subject to the virus. Thus, there exists a requirement to forecasting the possible spreading of virus within small districts. However, most present works are done only on a Marco-level failing to treat small districts as individual regions as a unit to produce forecast. Meanwhile, random events which almost have no effect on the prediction

made on Marco-level, may become an assignable element in small districts. Hence, to address the limitation, which is missed by present works, we propose an improved model which is still based on the traditional SEIRD model but introduce random walk to simulate the possible occurrence of random events. This implies our model will mainly focus on the handling of effects of random events during the application of SEIRD model, instead of improving a spatial-temporal related parameters which have almost been optimized by previous studies. Our detailed model will be explained in the following methodology part.

3. Methodology

Our model attempt to study and simulate how the covid-19 pandemic spread across Iowa. In order to achieve this goal, we separate our model by two-part. The first part of our model is the SEIRD model part, and the second part of our model is the random process part.

3.1. A SEIRD model

Firstly, let us come to the SEIRD model part, whose flow chart is given as Fig.1. To study how the covid-19 pandemic spread across Iowa, instead of studying Iowa as a whole, we divided Iowa into small county within Iowa. Then, we build the SEIRD model for each county inside Iowa. In SEIRD mode, S denotes susceptible, which means the group of people that could be infected by the diseases in the model. In our model, we assume that all population in a county is susceptible based on the fact that Covid-19 can infect almost everyone. Hence, the initial S is the entire population of the county in our model. E denotes exposed, which means the group of people who has exposed to the Covid-19 virus, in other words, the group of people exposed to Covid-19, and has a high chance to be infected. I denote infected, which means the group of people in the model who have already been infected by the disease. Then, in our SEIRD model, we assume these infected people will be converted to two group of people. One group is R, 'R' is the abbreviation of the removed, which allude to the group of the people which have recovered. Those recovered people will no longer be infected by the Covid-19 in our model since they will gain antibodies. Another group of people converted from I is D, which alludes to the group of people who died because of the infection of the covid-19.

In the first part of our SEIRD model, we suppose that the model is dynamics. Each group of people (S,E,I,R,D) is changing and converting to another group with time. We assume that the group of susceptible people in a county are converted to exposed people at a constant rate, which we assign Beta to present this constant value. Also, we assume that a group of exposed people is also converted to a group of infectious people at a constant rate within a county, and we assign sigma to represent this value. In addition, we assume the group of infected people is converted to a group of recovered people at a constant rate within a county, and we assign Gamma to represent this constant. Moreover, we assume the group of infectious people is converted to a group of dead people at a constant rate within a county, and we assign μ to represent this constant.

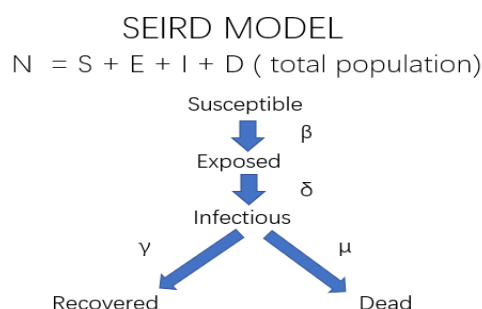


Figure 1. Flow chart of SEIRD MODEL

3.2. B Random walk

We did not stop here. In the second part of our model, we use a random process to build connections between each county within Iowa. In fact, people will not always stay in one county. They are actually moving across counties for commuting and traveling. Hence, we assume that the infected people in a county move in and move out in constants. To determine that the infected people in a county will move out to which nearby county, we run a random walk model in java to simulate this process to see the moving direction of the infected people.

In addition, it is clear that the constant in our model including β , σ , γ , μ will vary by their nature (due to the functional form error and real-world accident), and they will not be fixed by specific values all the time. Hence, we assume each of these constants follows a normal distribution. Therefore, in java, we use a random number generator to generate a random number for each of these constants each time following the probability density curve of normal distribution. Moreover, we note that the constants which represent the infected people in a county move in and move out should also not be fixed each time. Hence, we assume that they also follow normal distributions. So, we also use a random number generator to generate a random number for these constants each time following the probability density curve of normal distribution.

4. Experiments

4.1. Data source

We collect our data from counties the state of Iowa (of the U.S.) via the official website of the center for disease control and prevention and all data used in the simulation are available online. As Iowa is a state with a plain terrain in the central of the U.S. which means this region is hard to be affected by the exogenous infected and temporal inhomogeneities (e.g., mountains)

Meanwhile, we will assume fixed local total population numbers of each counties whose value will not change during the simulation. This method also been applied by another report aimed at identifying the SIERD model for COVID-19, as factors aside from COVID-19 pandemic and population mobility could also result in changes of total population [3]. Therefore, we will obey what the common literature will do to set the fixed total population of each county.

4.2. Simulation setting

In this work, all simulations are performed under java programming language with the JetBrains Integrated Development Environment. Then, the Lyon County of Iowa with a total population of 11,755 is selected to simulate the spreading of the virus. Additionally, our simulation will focus on the initial stage of the spreading so that the simulation results can hardly be interfered by irrelevant variables such as the vaccinated rate and block down policy [5]. Then, the regular flow rate of population rotation between Lyon and adjacent counties will be assumed to be 0.001 which is derived from a related work, as the specific rate of Lyon population flow has not been found [12]. Then, the basic parameter values of this SEIRD are assumed to be at a medium level with $\sigma = 1/4$, $\gamma = 1/10$, $R_0 = 3$, which is suitable for most regions within the US according to the study done by Korolev [3]. Eventually, Fig.2 as a flow chart is given here to help the understanding of the entire process of the simulation.

Note :
 $\sigma' = \sigma \times \text{basic rate} \times (1 + \text{fluctuation rate})$
 $\gamma' = \gamma \times \text{basic rate} \times (1 + \text{fluctuation rate})$
 $\beta' = \gamma' \times \text{basic reproduction number } R_0$
 Population flow in = Infected Population * flow rate * (1 + fluctuation rate)
 Population flow out = Infected Population * flow rate * (1 + fluctuation rate)
 Fluctuation rate -- generated by random walker individually for each equation

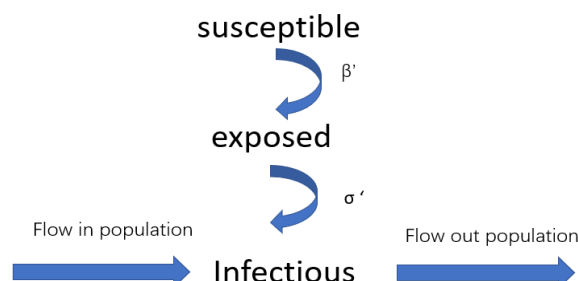


Figure 2. Flow chart of the entire simulation process

4.3. Simulation results

We use different pairs of SEIRD parameter values (σ and γ) in our model to simulate the spread of the pandemic since Aug 31, 2020. Compared with the real data represented by Fig.3, we found that, when applying the assumed basic parameter values ($\sigma = 1/4$, $\gamma = 1/10$, $R_0 = 3$), our model hardly to fit the line chart of real data, and the impact of fluctuation generated by random walk also could be ignored. However, when we apply the pairs of parameter values which is suitable for Japan ($\sigma = 1/5$, $\gamma = 1/18$, $R_0 = 1$), we found that our model is more likely to fit the line chart in Fig.2 within a month and the fluctuation generated by random walk also could impact the trend of spreading. The reason behind could be that the social distance of citizens in Iowa could be farer than other region of the US. Therefore, it is safe to conclude that our short-term forecasting model which considering random events maybe more suitable for regions with a low value of population density and basic reproduction number.

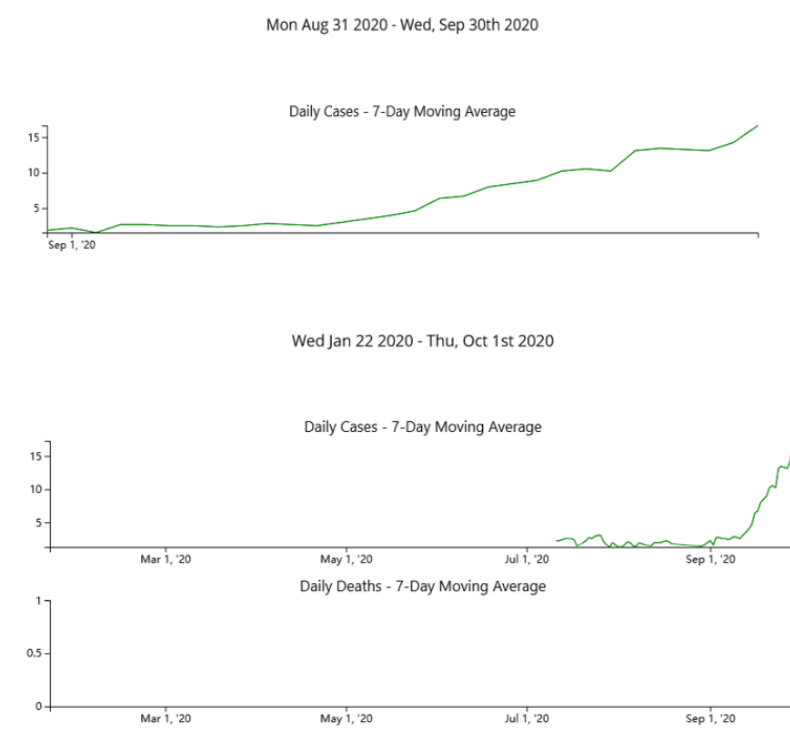


Figure 3. Time-series Chart of Cases in Lyon county

5. Discussion

This short-term prediction of small regions usually is adequate for the local authorities to distribution necessary medical supply and take correspond administering method to arresting pandemic.

However, this study has several limitations:

- (1) During the experiments, the random walk simulates the occurrence of random event based on whether the walker reaches the space boundary. However, the shape and the size of the boundary are not well-identified.
- (2) This model is still in an early version, this mean that these expected values of parameters now is constant values but still can be optimized as time-space functions in the further study.
- (3) While this model has successfully showed it is capable of complete short-term pandemic prediction in the county of Lyon, Iowa (with a level terrain), but this model is still suggested to be trained by data collected from regions with different kinds of terrains, so that this model could fit and be applied to more complex contexts.

6. Conclusion

In this paper we first overview the applications and main use of random walk and explore the present improvement on the SEIRD pandemic model. Afterwards, a prediction model of COVID-19 virus is proposed aimed at simulation the spreading of pandemic within a small region. In this new model, we manage to introduce random walk as a tool to simulate the actual value fluctuation around the expected valued of parameters in SEIRD pandemic model and the fluctuation of population mobility. Then we combine SEIRD model and random walk to establish the forecasting model which estimate the short-term pandemic situation within small districts.

Next, this model is evaluation by the public data of state of Iowa provided by the center for disease control and prevention (US) in a short term of 30-Days. The result shows this suggested model could fit the actual situation in a short-term but is hard to fit the long-run pandemic situation.

References

- [1] E. Karaçuha *et al.*, (2020) Modeling and Prediction of the Covid-19 Cases with Deep Assessment Methodology and Fractional Calculus, In: *IEEE Access*, vol. 8, pp. 164012-164034.
- [2] Z. Liu, J. Zuo, R. Lv, S. Liu and W. Wang, (2021) Coronavirus Epidemic (COVID-19) Prediction and Trend Analysis Based on Time Series, In: *IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)*, pp. 35-38.
- [3] Korolev, Ivan. (2021) Identification and estimation of the SEIRD epidemic model for COVID-19. *Journal of econometrics*, 220,1: 63-85.
- [4] Viguerie, Alex, et al. (2021) Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion. *Applied Mathematics Letters* 111: 106617.
- [5] K. Tang, Y. Ma, H. Yan and Y. Rao, (2021) A New Prediction Method of COVID-19 Epidemic, In: *2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pp. 331-335.
- [6] Gibson, Graham C., Nicholas G. Reich, and Daniel Sheldon, (2020) Real-time mechanistic bayesian forecasts of covid-19 mortality. *MedRxiv*.
- [7] F. Xia, J. Liu, H. Nie, Y. Fu, L. Wan and X. Kong, (2020) Random Walks: A Review of Algorithms and Applications, In: *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 2, pp. 95-107.
- [8] S. Yan and G. Wu, (2020) Fitting of Impact of COVID-19 Pandemic on S&P 500 Index Using Random Walk, In: *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, pp. 1007-1011.

- [9] Quintero, Yullis et al. (2021) Machine learning models for the prediction of the SEIRD variables for the COVID-19 pandemic based on a deep dependence analysis of variables. *Computers in biology and medicine*, 134: 104500.
- [10] W. Douhan and X. Qinneng, (2020) Economic Evaluation of Quarantine: A Case Study of COVID-19 Pandemic in Belgium, In: *2020 5th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pp. 99-102.
- [11] M. Naveed, S. Qazi and B. A. Khawaja, (2021) UAV-based Life-Saving Solution for Police to Maintain Social-Distancing During Covid-19 Pandemic Using 4G-LTE Technology," In: *2021 International Conference on Communication Technologies (ComTech)*, pp. 28-32.
- [12] S. Zhang, (2020) Analysis of Blocking Population Flow to Control COVID-19 in Multi-regions Based on Discrete SEIR Epidemic Model, In: *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*, pp.147-152.