Estimating the Optimal Parameters of SEIR model for Predicting the Trajectory of COVID-19 Pandemic in Japan

1st Taichi Fukuta*
Nagoya Institute of Technology
Nagoya, Japan
t.fukuta.664@stn.nitech.ac.jp

2rd Hirotaka Ito
Nagoya Institute of Technology
Nagoya, Japan
ht-itoh@nitech.ac.jp

3nd Kenji Funahashi
Nagoya Institute of Technology
Nagoya, Japan
kenji@nitech.ac.jp

2021 IEEE 10th Global Conference on Consumer Electronics (GCCE) | 978-1-6654-3676-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GCCE53005.2021.9622073

***Abstract*—A system that uses the susceptible-exposed-infectious-removed (SEIR) model and some methods to predict COVID-19 (SARS-Cov-2) trajectory of the pandemic which is currently prevalent. This paper reports on how to estimate the optimal parameters of the SEIR model, which can predict the pandemics of infectious diseases.**

***Index Terms*—COVID-19, SEIR Model, Grid-Search, Holt-Winters Method.**

I. Introduction

As of August 2021, COVID-19 (SARS-Cov-2) is still prevalent. In Japan, despite the implementation of various policies and scientific measures, the COVID-19 pandemic has not converged. Therefore, to at least predict the final period of this pandemic, a system was developed that predict it using the susceptible-exposed-infectious-recovered (SEIR) model [1], which is effective for predicting pandemics of infectious diseases. Moreover there exits studies that simulated the trajectory of COVID-19 pandemic at each country [2] [3] [4] [5]. Section II presents the introduction about related works. Section III presents the explanation of the SEIR model. Section IV presents an explanation on how to estimate paramters of SEIR model and how to predict COVID-19 pandemic. Section V reports the result of predicting by the above mentioned system. Therefore, it can be said that this study is sufficiently novel compared to other studies using the SEIR model.

II. Related Works

One of the related works is a simulation of the COVID-19 epidemic that began in China in the winter of 2020 [2]. As a result, it was considered from the parameters of the optimized SEIR model that Wuhan was the most active virus. The next work is a study that also simulated an epidemic in India at the spring of 2020 [3]. Here, both the SEIR model and the linear regression model were used, and it was said that the former model was more accurate as a result of the experiment. It was also said that the SEIR model obtained a certain degree of accuracy for experiments simulating pandemics in Spain and Italy [4], Pakistan and Malaysia [5]. However, these works only verified the accuracy of the known past data and the calculation results of the SEIR model, and did not predict the future pandemic, which is unknown data. In that regard, the purpose of this study is to infer the parameters of the SEIR model required to predict the future COVID-19 pandemic from past simulation results and to verify how much performance they bring to the prediction. Therefore, it can be said that this study is sufficiently novel compared to other studies using the SEIR model.

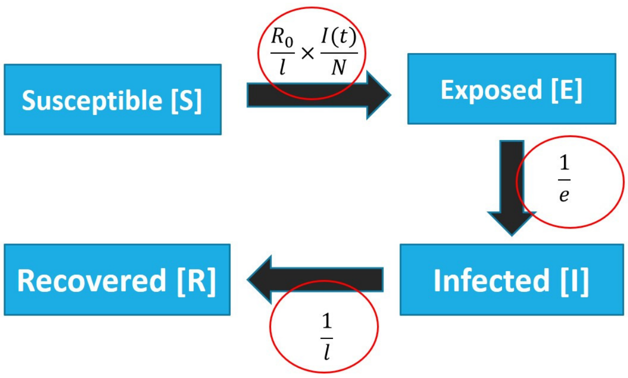

Fig. 1. Structure of SEIR model

III. SEIR Model

When an infectious disease cause a pandemic, the country's population is divided into four categories: (1) the population that is at high risk of inflection (susceptible), (2) the population that has not yet developed symptoms after infection (exposed), (3) the population with symptoms (infected), and (4) the population that has recovered or died after infection (recovered). The model that predict these categories is the "SEIR" model. It predicts the changes in these populations per day (t) by formulating the following ordinary differential equations:

$$\frac{d}{dt}S(t) = -\frac{R_0 I(t)}{eN}S(t) \quad (1)$$

$$\frac{d}{dt}E(t) = \frac{R_0 I(t)}{eN}S(t) - \frac{E(t)}{e} \quad (2)$$

$$\frac{d}{dt}I(t) = \frac{E(t)}{e} - \frac{I(t)}{l} \quad (3)$$

$$\frac{d}{dt}R(t) = \frac{I(t)}{l} \quad (4)$$

$$N : total\ \ population$$

The expression of structure of SEIR model is show in Fig. 1. R_0 is the "effective reproductive number" which shows how many people can be infected by one infected person, e is the waiting time from infection to onset, and l is the waiting time from onset to recovery or death. Also, $\frac{R_0 I(t)}{eN}$ represents the rate at which the population contained in "susceptible" transitions to "exposed", $\frac{1}{e}$ represents the rate at which the population contained in "exposed" transitions to "infected", $\frac{1}{l}$ represents the rate at which the population contained in "infected" transitions to "Recovere" each the time elapses. In addition because $N = S(t) + E(t) + I(t) + R(t)$, and (1) – (4) satisfies $(1) + (2) + (3) + (4) = 0$, the population forecast by the SEIR model assumes that the total population of the area (N) does not change with time.

IV. How to predict COVID-19 pandemic

The following three steps are taken to predict the COVID-19 trajecty using the SEIR model.

Step A: Estimate the past optimal parameters

The SEIR model is trained for past data on four kinds of population in Japan to find the parameters which minimize the root-mean-squared-error (RMSE) by "Grid-Search". RMSE is derived as follows:

$$RMSE(i) = \sqrt{\frac{1}{n}\sum_{t=i}^{n}(y_t - \hat{y}_t)^2}$$

y_t :measured value, $\hat{y}_t$:predicted value

Here, the RMSE between the measured values for "susceptible" and "recovered" that are currently clear is evaluated, and these predict values for the output of the SEIR model. Subsequently, if the output of the SEIR model is for "S" days , length of past data is "T" and prediction start date is "i", the optimal parameter P_i is as follows:

$$P_i = \underset{R_0,e,l}{\arg\min}[RMSE^{(Susceptible)}(i) + RMSE^{(Recovered)}(i)]$$

In addition, the past optimal parameter P_{past} is as follows:

$$P_{past} = (P_0, P_1, ..., P_{T-S})$$

So, the algorithm that derive P_{past} is shown in Fig. 2.

Step B: Estimate the future optimal parameters

In previous studies [4] [5], parameters were inferred for known past epidemic data using a method similar to $Step\ A$. However, in this study, the future optimal parameters (P_{future}) is estimated from P_{past} obtained in $Step\ A$ to predict an unknown future pandemic. Therefore, the "Holt-Winters Method" [6] which is a time series analysis is used.

Require: length of past data T, length of output S groups of parameter, R_0, e, l

for t = 0 to $T - S$ **do**
  best_score $\leftarrow 1.0 * 10^8$
  {Grid-Search start}
  for $i = a \in R_0$ **do**
    for $j = b \in e$ **do**
      for $k = c \in l$ **do**
        $s \leftarrow RMSE^{(susceptible)}$
        $r \leftarrow RMSE^{(recovered)}$
        if $(s + r) \leq$ best_score **then**
          best_score $\leftarrow s + r$
          $P_t \leftarrow (i, j, k)$
        end if
      end for
    end for
  end for
  {Grid-Search end}
end for
return $P_{past} = (P_0, P_1, ..., P_{T-S})$

Fig. 2. Pseudo code of algorithm for estimating P_{past}

1) Holt-Winters Method: The method uses three elements of time series data, a value that fluctuates randomly without depending on time series ($level$), long-term inclination of time series data ($trend$), and the effect that the value repeats at regular intervals ($season$). Using the time series value y_t, the future value $\hat{y}_{t+1}$ is expressed as follows:

$$\hat{y}_{t+1} = level_{y_t} + trend_{y_t} + season_{y_t}$$

As an example, the result of the applying "Holt-Winters Method" to the quarterly UK gas consumption "UKgas" [7] which is the standard dataset in library of programing language R is shown in Fig. 3. Focusing on $level$ and $trend$ of measured value, it can be seen that these are on an increasing trend overall. This may mean that the demand for gas is gradually

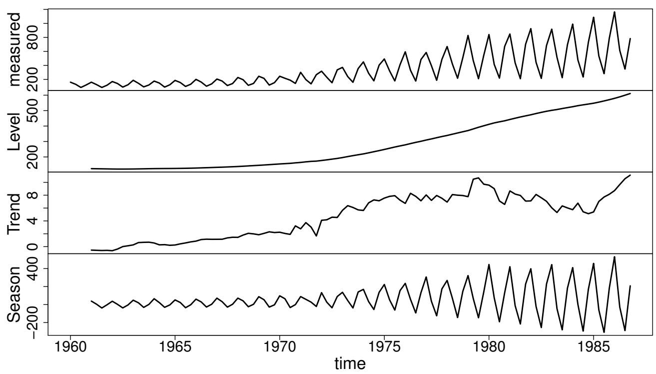

Fig. 3. Decompose of quarterly UK gas consumption (UKgas)

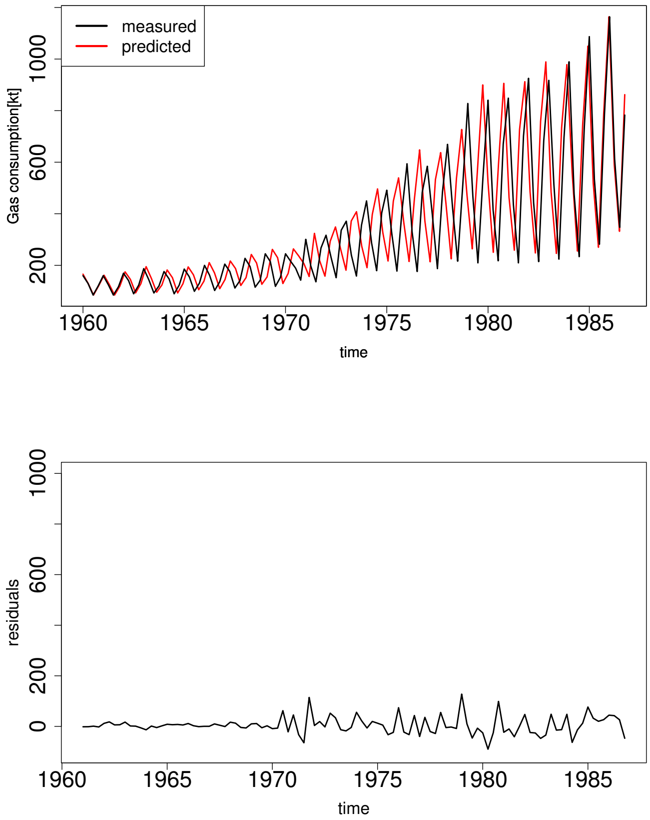

Fig. 4. Predicted value, measured value of UKgas and residuals

increasing over time. For *season* of measured value, the amplitude increases over time. This indicates that there is a large difference in gas consumption in each season as the years go by. Especially in winter, it is thought that the difference in consumption from other seasons is remarkable due to the fact that a lot of energy is consumed to keep warm and the demand for gas is increasing.

Next, the measured value of gas consumption , the predicted value of gas consumption by the "Holt-Winters Method", residuals between them are shown in Fig. 4. $RMSE$ of gas consumption was 34.66. Therefore, in this case, it can be said that there is no big difference between the predicted value by "Holt-Wintes Method" and the measured value.

2) How to estimate future optimal parametrs by "Holt-Winters Method": Let the P'_{past} optimal parameters from $T-S$ to $T+1$, P'_{past} is as follows:

$$\begin{aligned} P_{T-S+1} &= level_{P_{T-S}} + trend_{P_{T-S}} \\ &\quad + season_{P_{T-S}} \\ &\vdots \\ P_{T+1} &= level_{P_T} + trend_{P_T} + season_{P_T} \\ P'_{past} &= (P_{T-S+1}, P_{T-S+2}, ..., P_{T+1}) \end{aligned}$$

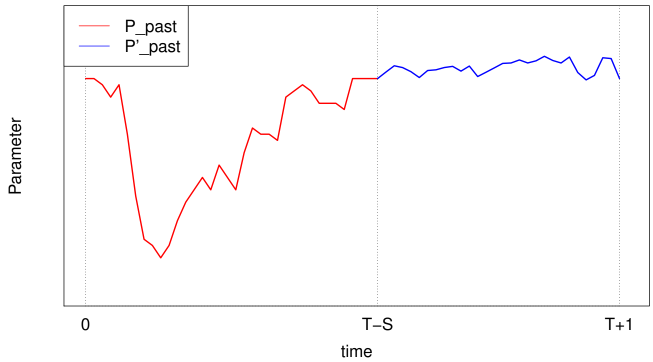

Fig. 5. Example of estimated P_{past} and P'_{past}

The example of estimated P_{past} and P'_{past} is shown in Fig. 5. Here, the optimal parameters used when predicting future pandemics are those at time $T+1$. So, P_{future} is as follows:

$$P_{future} = P_{T+1}$$

Step C: Input SEIR model the future optimal parameters

Finally, inputting the P_{future} derived in *Step B* into the SEIR model completes the prediction of future pandemic.

V. Verification of Predictioin Accuracy

A. Experiment

The purpose of experiment is to infer P_{future} required to predict the future COVID-19 pandemic from P_{past} and to verify how much performance they bring to the prediction. Setting of *Step A* and *Step C* is shown in TABLE I. This time, the focus is on the length of output in *Step A*, and the difference in accuracy for each is verified. So, there are 5 kind of lengths of output (days) in *Step A*: 7, 14, 21, 28, 35. Period of past data is set from 07/01/2020 to 02/28/2021 and period of future prediction is set from 03/01/2021 to 03/31/2021. Here, MAPE is used as an evaluation index. MAPE is derived as follows:

$$MAPE = \frac{1}{n}\sum_{t=0}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

y_t :measured value, $\hat{y}_t$:predicted value

Measured value of "susceptible" uses a minus the infected persons number of cumulative from there the total population of Japan [8] as 120 million so far and measured value of "recovered" uses a total of recoverers and dead in Japan [8] for calucurating RMSE and MAPE. Regarding the experimental environment, the programming language was R, and the processor was Intel (R) Core (TM) i5-10300H CPU @ 2.50GHz

TABLE I
SETTING OF *Step A* AND *Step C*

Step A	Period of past data	07/01/2020 – 02/28/2021
	Length of output (days)	7, 14, 21, 28, 35
	Range of parameters	$0.1 \leq R_0 \leq 3.0$
		$5.0 \leq e \leq 10.0$
		$6.0 \leq l \leq 14.0$
Step C	Period of future prediction	03/01/2021 – 03/31/2021

TABLE II
FUTURE OPTIMAL PARAMETERS AND MAPE OF PREDICTION RESULT

Length of output in *Step A*	R_0	e	l	MAPE	
				susceptible	recovered
7 days	2.09	10.28	10.57	2.71e-04	2.35-e02
14 days	3.18	10.00	14.00	5.54e-04	2.67-e02
21 days	2.51	10.00	15.19	2.52e-04	1.27-e02
28 days	1.08	10.00	13.96	7.00e-05	5.47-e03
35 days	1.13	10.00	13.96	6.03e-05	1.74-e03

B. Result

The future optimal parameters obtained in *Step B* and MAPE of prediction result are shown in TABLE II. and the prediction results of "susceptible" and "recovered" are shown in Fig. 6 and Fig. 7. In the prediction for a total of 31 days from March 1 to March 31, 2021, the forecast results for "susceptible" and "recovered" were close to the measured values when the output length in *Step A* was 28 or 35 days. In *Step A*, the P_{past} was estimated to be the parameter with the smallest error when predicting the past pandemic for the length of output in *Stap A* by SEIR model. It is considered that the nature of P_{past} were not lost in the process of sequentially deriving P_{past} to P_{future} by the "Holt-Winters Method". Therefore, it was assumed that if the length of the output in *Step A* is set close to that in *Step C*, the prediction accuracy of this system would be high.

VI. CONCLUSION

In this paper, we suggested how to estimate the optimal parameters of the SEIR model, and predicted the trajectory of the COVID-19 pandemic prevalent during March 2021 in Japan. In future works, we plan on verifying the prediction accuracy when estimating parameters by a time series analysis method other than the Holt-Winters method, and the generalization performance in a prediction period other than March 2021.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP19K03028.

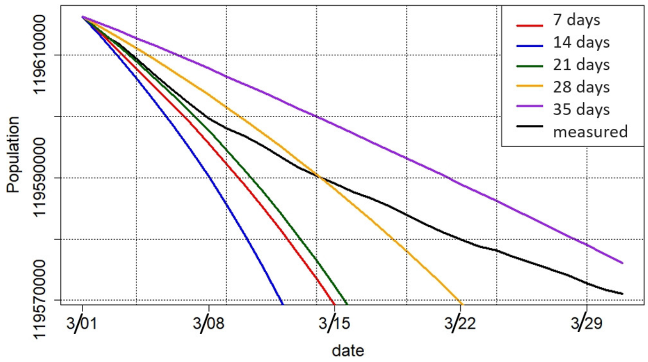

Fig. 6. Difference in prediction of "susceptible" in output length in *Step A*

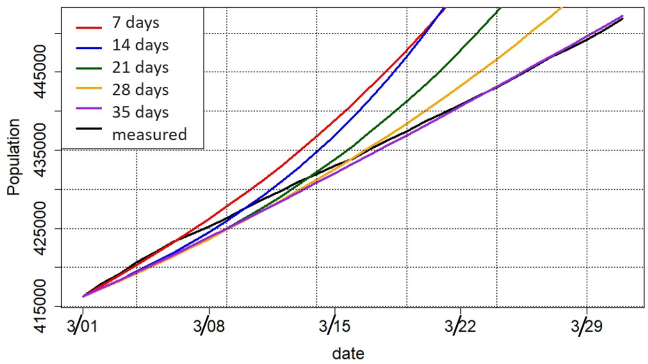

Fig. 7. Differences in prediction of "recovered" in output length in *Step A*

REFERENCES

[1] H. Nishiura, and H. Inaba, "Prediction of infectious disease pandemics: Quantitative Challenges in Mathematical Models of Infectious Diseases" ("Kansensyou ryuukou no yosoku : Kansensyou suuri moderu ni okeru teiryouteki kadai"), Statistical mathematics (Tokei Suuri), Japan, vol. 54, No. 2, pp. 461–480, 2006

[2] L. Peng, W. Yang, T. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling", arXiv preprint arXiv, February 2020.

[3] G. Pandey, P. Chaudhary, R. Gupta, and S. Pal, "SEIR and Regression Model based COVID-19 outbreak predictions in India", Pandey et al, 2020.

[4] L. Lopez, and X. Rodo, "A modified SEIR model to predict the COVID-10 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemic", Results in Physics 21, No.103764, 2021.

[5] H. Alrabaiah, M. Afran, K. Shah, Ibrahim Mahariq, and A. Ukkah, "A comprative study of spreading of novel corona virus disease by ussing fractional order modified SEIR model", Alexandra Engineearing Journal (2021), vol. 60, pp. 573–585, 2021.

[6] P. Goodwin. "The Holt-Winters approach to expotential smoothing: 50 years old and going strong" , The International Journal of Applied Forecasting, Vol. 19, No. Fall, pp. 30-33, 2010.

[7] J. Durbin, and S. J. Koopman, "Time Series Analysis by State Space Methods", Oxford University, 2001.

[8] Coronavirus Disease (COVID-19) Situation Report in Japan, https://toyokeizai.net/sp/visual/tko/covid19/en.html