The daily computed weighted averaging basic reproduction number $R_{0,k,\omega}^n$ for MERS-CoV in South Korea

Darae Jeong [a], Chang Hyeong Lee [b], Yongho Choi [a], Junseok Kim [a,*]

[a] Department of Mathematics, Korea University, Seoul 136-713, Republic of Korea
[b] Department of Mathematical Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 689-798, Republic of Korea

HIGHLIGHTS

- We propose the daily basic reproduction number for MERS-CoV in South Korea.
- We use an SIR model with piecewise constant contact and removed rates.
- We apply the explicit Euler's method for the solution of the SIR model.
- We use nonlinear least-square fitting procedure for finding the best parameters.

ARTICLE INFO

ABSTRACT

In this paper, we propose the daily computed weighted averaging basic reproduction number $R_{0,k,\omega}^n$ for Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, May to July 2015. We use an SIR model with piecewise constant parameters β (contact rate) and γ (removed rate). We use the explicit Euler's method for the solution of the SIR model and a nonlinear least-square fitting procedure for finding the best parameters. In $R_{0,k,\omega}^n$, the parameters n, k, and w denote days from a reference date, the number of days in averaging, and a weighting factor, respectively. We perform a series of numerical experiments and compare the results with the real-world data. In particular, using the predicted reproduction number based on the previous two consecutive reproduction numbers, we can predict the future behavior of the reproduction number.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Middle East respiratory syndrome coronavirus (MERS-CoV), first identified in Saudi Arabia in June 2012, is a viral respiratory disease caused by a novel coronavirus [1]. South Korea experienced the largest outbreak of MERS-CoV infections outside the Arabian peninsula [2]. There have been 186 confirmed infective cases in South Korea within two months after the first infective person who returned from a trip to the Arabian peninsula was diagnosed on 11 May, 2015, and 38 people died and more than 16,000 people were quarantined due to the spread of the disease. Fig. 1 represents the epidemic curve of MERS-CoV, South Korea, 20 May–17 July 2015 [3,4].

In order to implement appropriate surveillance and control measures, it is very important to predict the future trend of the epidemic. Therefore, it is the purpose of the present paper to show daily computed reproduction numbers for epidemics, in particular, MERS-CoV, so that we can predict the future behavior of the reproduction number day-by-day.

---

* Corresponding author.
E-mail address: cfdkim@korea.ac.kr (J. Kim).
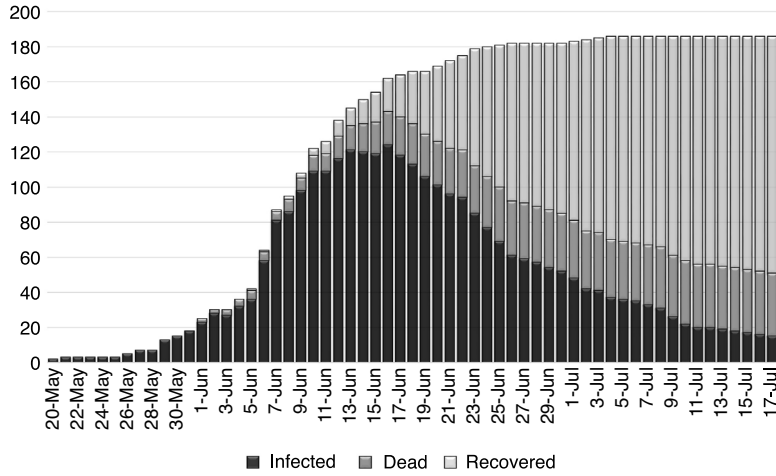*URL:* http://math.korea.ac.kr/~cfdkim/ (J. Kim).

Fig. 1. Epidemic curve of MERS-CoV, South Korea, 20 May–17 July 2015.

In the mathematical modeling of the spread of an epidemic disease, it is crucial to estimate the parameters (e.g., contact rate and recovery rate in an SIR model), but it is difficult to do so due to the lack of data available for the estimation. In case of the MERS-CoV spread in South Korea, there are almost complete daily data of the spread of the disease, and we choose the MERS-CoV spread case in South Korea as an epidemic model to verify the validity and accuracy of the method proposed in this paper. To the authors' knowledge, there have been no previous works that present computational methods for estimation of the values of the epidemic parameters based on actual daily epidemic data.

The rest of the paper is organized as follows. In Section 2, we describe the mathematical model. In Section 3, we provide a numerical algorithm for the estimation of the parameters. We perform several numerical experiments in Section 4. In Section 5, we provide a summary and present our conclusions.

2. The mathematical model

In this paper, we consider the SIR model that was introduced in 1927 by A.G. McKendrick and W.O. Kermack [5]. The model is simple and has been widely used so far, for instance, in multigroup epidemic modeling [6], online social network dynamics [7], the model adopting the delay term [8], stochastic model [6,9,10], vaccination strategy [11,12]. In this model, the population is divided into susceptible $S(t)$, infected $I(t)$, and removed $R(t)$ individuals at time t. The governing ordinary differential equations for the SIR model are as follows:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t), \tag{1}$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t), \tag{2}$$

$$\frac{dR(t)}{dt} = \gamma I(t) \tag{3}$$

with initial condition $S(0) = S^0, I(0) = I^0$, and $R(0) = R^0$. Here, the parameters β and γ denote the contact (susceptibility to disease) and removed (either dead or recovered) rates from disease, respectively [13]. We assume that a removed individual can never be infected again. Let N be the constant total population size. Therefore, it is sufficient to solve only two Eqs. (1) and (2), i.e., $R(t) = N - S(t) - I(t)$. By assuming that β and γ are time-dependent parameters, we generalize Eqs. (1)–(3) as follows:

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t), \tag{4}$$

$$\frac{dI(t)}{dt} = \beta(t)S(t)I(t) - \gamma(t)I(t), \tag{5}$$

$$\frac{dR(t)}{dt} = \gamma(t)I(t). \tag{6}$$

For example, the transmission of the vector-borne diseases such as Dengue fever and Malaria has strong seasonal patterns and thus the parameters are estimated as the time-dependent functions [14,15]. Another example is a seasonal SIR model for the spread of the whooping cough in which the contact rate β is given as a time-dependent periodic function which accounts seasonal changes [16].
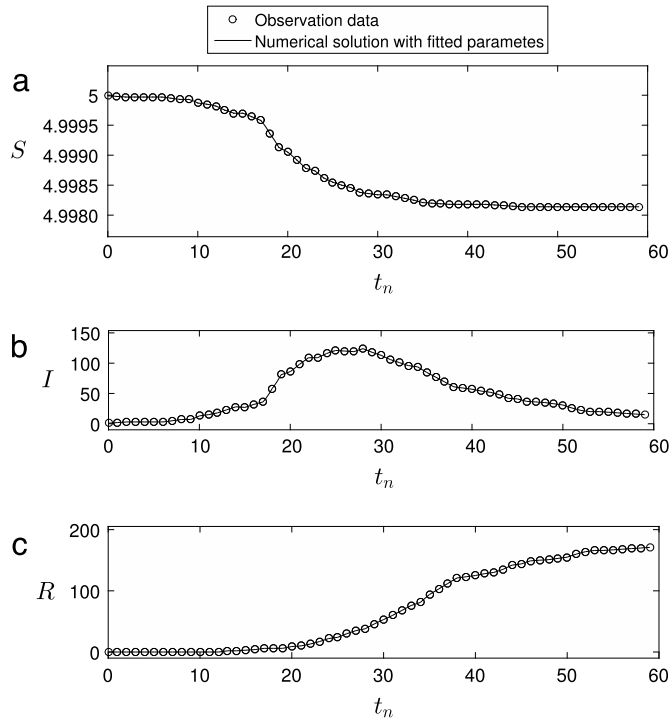
Fig. 2. Time series of observed MERS-CoV epidemiological data and numerical solutions of (a) S, (b) I, and (c) R, respectively. Here, $N = 500{,}000$ is used.

3. The numerical method

Let us assume that we have daily statistics about S, I, and R. Let $t_n = t_{\text{ref}} + n$ be time, where t_{ref} is a reference time. Let DS^n, DI^n, and DR^n be observed susceptible, infected, and removed data at time $t = t_n$, respectively. Let S^n, I^n, and R^n be numerical approximations of $S(t_n)$, $I(t_n)$, and $R(t_n)$, respectively. In this paper, we propose a computation of $\beta(t)$ and $\gamma(t)$ on a daily basis, which fit best to the real data. To be more specific, for given data (DS^n, DI^n, DR^n) and $(DS^{n+1}, DI^{n+1}, DR^{n+1})$, we want to find piecewise constant parameters β^{n+1} and γ^{n+1} which minimize the following

$$\min_{\beta^{n+1}, \gamma^{n+1}} \left[(S^{n+1} - DS^{n+1})^2 + (I^{n+1} - DI^{n+1})^2 + (R^{n+1} - DR^{n+1})^2 \right], \tag{7}$$

where S^{n+1}, I^{n+1}, and R^{n+1} are the numerical solutions of Eqs. (4)–(6) with initial condition $(S^n, I^n, R^n) = (DS^n, DI^n, DR^n)$. We divide one day between t_n and t_{n+1} into m subintervals, then $\Delta t = 1/m$ is the time step size. Let $S^{n+1,1} = S^n$ and $I^{n+1,1} = I^n$, then by applying the explicit Euler's method to Eqs. (4) and (5), we have

$$S^{n+1,k+1} = S^{n+1,k} - \beta^{n+1} S^{n+1,k} I^{n+1,k} \Delta t, \tag{8}$$

$$I^{n+1,k+1} = I^{n+1,k} + \left(\beta^{n+1} S^{n+1,k} I^{n+1,k} - \gamma^{n+1} I^{n+1,k} \right) \Delta t, \tag{9}$$

$$R^{n+1,k+1} = N - S^{n+1,k+1} - I^{n+1,k+1}, \quad \text{for } k = 1, \ldots, m, \tag{10}$$

where we have used the condition $R(t) = N - S(t) - I(t)$. Unless otherwise stated, we use $\Delta t = 1/100$ in all numerical tests. If the solutions of Eqs. (8)–(10) are obtained for all k, then we define $S^{n+1} = S^{n+1,m+1}$, $I^{n+1} = I^{n+1,m+1}$, and $R^{n+1} = R^{n+1,m+1}$. Note that while we can use higher-order numerical methods such as Runge–Kutta schemes to solve Eqs. (4) and (5), we use the explicit Euler's method for the sake of simplicity. Next, to find the best β^{n+1} and γ^{n+1}, we use a MATLAB routine called **lsqcurvefit**, based on the least-squares sense, which finds coefficients β^{n+1} and γ^{n+1} that solve the problem (7). We use $\beta^{n+1} = 1.0e-6$ and $\gamma^{n+1} = 1.0e-1$ as an initial guess in all numerical simulations. The MATLAB code is given in Appendix for the interested readers.

4. Numerical simulations

In all computations, the initial conditions are taken to be $N = 500{,}000$, $S^0 = N - 1$, $I^0 = 1$, and $R^0 = 0$ with $t_{\text{ref}} = 20$ May 2015. Fig. 2 shows the time series of observed MERS-CoV epidemiological data and numerical solutions S, I, and R, respectively.

Table 1
Error which is expressed by $E^n = R^n_{linear} - R^n_{0,1,0}$ from 18 June, 2015 to 17 July, 2015, i.e., $n = 30, \ldots, 59$.

n	E^n	n	E^n	n	E^n
30	−2.4532	40	−0.1110	50	−0.0002
31	+0.3215	41	+0.0000	51	+0.0001
32	−0.6608	42	−0.0001	52	−0.0001
33	+0.3751	43	−0.1999	53	−0.0002
34	−0.2251	44	+0.2571	54	−1.6766
35	+0.5175	45	−0.4144	55	+3.3534
36	−0.0958	46	+0.6574	56	−1.6766
37	−0.1967	47	−0.1004	57	−0.0000
38	−0.0000	48	−0.1997	58	+0.0009
39	+0.1110	49	+0.0002	59	−0.0000

The basic reproduction number R_0 is the average number of secondary infections caused by one infectious individual in a completely susceptible population. It is an important measurement in that one can determine the stability of the disease-free equilibrium by computing the value of R_0; $R_0 < 1$ implies that the disease-free equilibrium is locally asymptotically stable [17], and if $R_0 > 1$, it is unstable [18,19]. Finding basic reduction number R_0 can be done through the next generation matrix [20,21] whose spectral radius is defined as R_0. In the SIR model, the basic reproduction number is computed as $R_0 = \beta N / \gamma$ [22,23], which will be used for our method in the next section.

4.1. Weighted averaging basic reproduction number $R^n_{0,k,\omega}$

In this section, we propose the weighted average values for infected and removed rates, and then define weighted averaging basic reproduction number. Let $\beta^n_{1,1} = \beta^n$ for $n = 1, \ldots, m$. We define k-days ($k \geq 2$) weighted average $\beta^n_{k,\omega}$ as

$$\beta^n_{k,\omega} = \begin{cases} \displaystyle\sum_{i=1}^{k} (\omega^{i-1} \beta^{n-i+1}) \Big/ \sum_{i=1}^{k} \omega^{i-1} & \text{if } n \geq k, \\ \displaystyle\sum_{i=1}^{n} (\omega^{i-1} \beta^{n-i+1}) \Big/ \sum_{i=1}^{n} \omega^{i-1} & \text{if } n < k, \end{cases} \tag{11}$$

where $0 < \omega \leq 1$ and the superscript $i - 1$ for ω^{i-1} represents an exponent. Also, $\gamma^n_{k,\omega}$ is defined similarly as Eq. (11). Finally, we define the daily computed weighted averaging basic reproduction number as $R^n_{0,k,\omega} = \beta^n_{k,\omega} N / \gamma^n_{k,\omega}$. Fig. 3(a)–(c) represent $\beta^n_{k,\omega}$, $\gamma^n_{k,\omega}$, and $R^n_{0,k,\omega}$ for $k = 5$, respectively, with $w = 0, 0.3$, and 1. Fig. 4 shows the results of the parameters for $k = 10$. Note that when $\omega = 0$, we have $R^n_{0,k,0} = R^n_{0,1,0}$ for all k. If k is large, then we have more smooth profile and the transition point ($R^n_{0,k,\omega} = 1$) moves to the right. If ω is close to zero, then we have a similar profile to $R^n_{0,1,0}$. Note that we have extraordinarily large values of $R^n_{0,k,\omega}$ from $n = 1$ to approximately $n = 20$ since the recovered individuals were almost zero ($\gamma^n_{k,\omega} \approx 0$) in early times and the infected ones increased rapidly.

4.2. Predictability of a linear extrapolation, R^n_{linear}

In this section, we investigate the predictability of a linear extrapolation of $R^n_{0,1,0}$. Let us define the linear extrapolation

$$R^n_{linear} = 2R^{n-1}_{0,1,0} - R^{n-2}_{0,1,0}, \quad \text{for } 3 \leq n \leq m, \tag{12}$$

where R^n_{linear} means the predicted reproduction number based on the previous two consecutive reproduction numbers and the error is defined as

$$E^n = R^n_{linear} - R^n_{0,1,0}, \quad \text{for } 3 \leq n \leq m. \tag{13}$$

Fig. 5(a) shows R^n_{linear} and $R^n_{0,1,0}$. Up to about three weeks, there are large differences between two reproduction numbers. However, we can observe a good agreement interval from $n = 30$ to $n = 59$, see Fig. 5(b). Fig. 5(c) and (d) are corresponding errors E^n of (a) and (b). The errors E^n from $n = 30$ to $n = 59$ are listed in Table 1. Except some days, the absolute errors $|E^n|$ are less than one. We note that we have a spike in Fig. 5(b) and (d). That is due to no population data changes for two days, i.e., 11 July and 12 July as shown in Fig. 1.

4.3. The SIR model with disease-related death

In this section, we consider the modified SIR model with disease-related death. Let the total population N be divided into susceptible $S(t)$, infected $I(t)$, recovered $R(t)$, and dead $D(t)$ individuals at time t, then the SIR model with fatality rate is
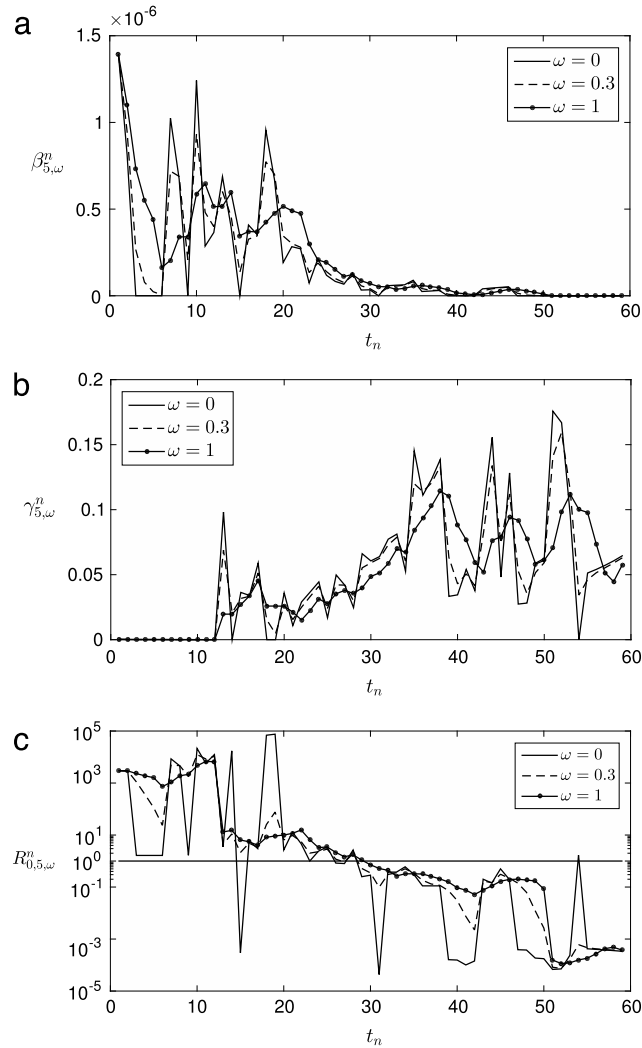
Fig. 3. Weighted averaging $\beta_{k,\omega}^n$, $\gamma_{k,\omega}^n$, and $R_{0,k,\omega}^n$ for $k = 5$. In each case, we compare the values with $\omega = 0, 0.3$, and 1.

given by

$$\frac{dS(t)}{dt} = -\beta(t)S(t)I(t), \tag{14}$$

$$\frac{dI(t)}{dt} = \beta(t)S(t)I(t) - \gamma(t)I(t) - d(t)I(t), \tag{15}$$

$$\frac{dR(t)}{dt} = \gamma(t)I(t), \tag{16}$$

$$\frac{dD(t)}{dt} = d(t)I(t) \tag{17}$$

with initial condition $S(0) = S^0$, $I(0) = I^0$, $R(0) = R^0$, and $D(0) = D^0$. Here, the time-dependent parameters $\beta(t)$, $\gamma(t)$, and $d(t)$ denote the contact, recovered, and dead rates from disease, respectively [22]. Now, by using the proposed method, we evaluate the numerical solution of the modified SIR model (14)–(17). We take the same initial conditions: $N = 500,000$, $S^0 = N - 1$, $I^0 = 1$, $R^0 = 0$, and $D^0 = 0$. The initial guesses for finding coefficients β, γ, and d are $1.0e{-}6$, $1.0e{-}1$, and $1.0e{-}1$, respectively. Using this model, we can predict the ratio of individuals who died by disease to the infected population at a given time. To investigate the ratio, we introduce the fatality percentage rate μ^n which is defined as

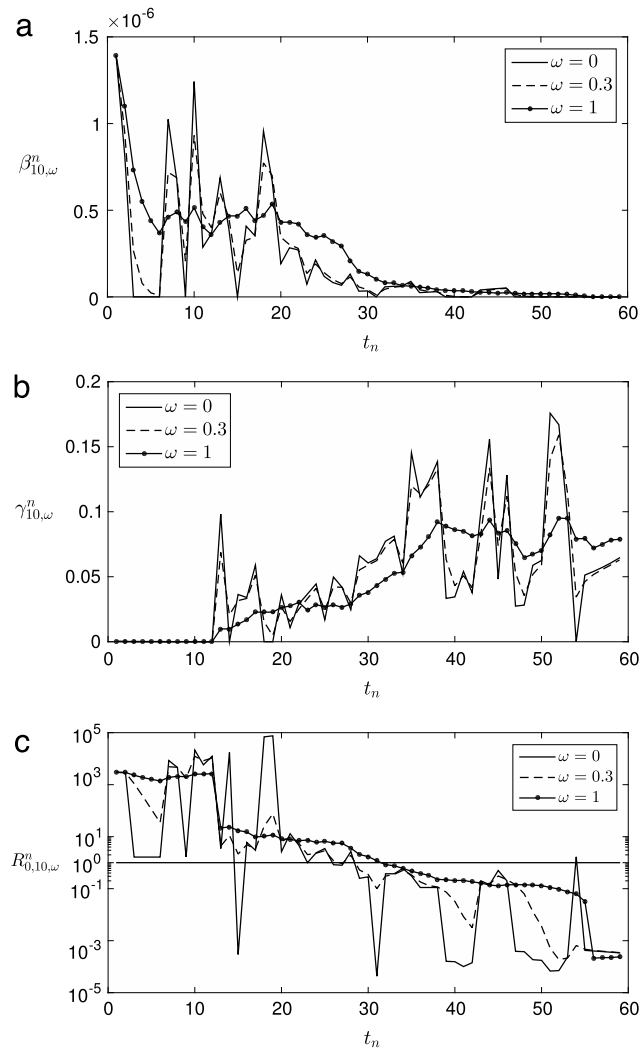$$\mu^n = \frac{D^n}{I^n + R^n + D^n} \times 100 \ (\%).$$

Fig. 4. Weighted averaging $\beta_{k,\omega}^n$, $\gamma_{k,\omega}^n$, and $R_{0,k,\omega}^n$ for $k = 10$. In each case, we compare the values with $\omega = 0, 0.3$, and 1.

In Fig. 6, we show the fatality percentage rate μ^n from the real observation data and the numerical approximations. Here, the fatality percentage rate μ^n from the numerical approximation is obtained by the numerical solutions S^n, I^n, R^n, and D^n. The numerical solutions S^n, I^n, R^n, and D^n are evaluated by using the best-fitting parameters β^{n-1}, γ^{n-1}, and d^{n-1} at time t_{n-1}. As shown in Fig. 6, the both results are in good agreement in later times.

5. Conclusion

We have proposed the daily computed weighted averaging basic reproduction number $R_{0,k,\omega}^n$ for MERS-CoV outbreak in South Korea, May to July 2015. The linearly extrapolated reproduction number, R_{linear}^n, has important implications for the future prediction of the trend of the reproduction numbers. We showed that the results computed by our method match very well with the actual data of the MERS-CoV spread in South Korea. Using the method proposed in this paper, one can predict the progress of the spread of infections on the use of real-time epidemic data, so that proper control policies can be performed in an appropriate time for reducing the spread of the infections. As future research, the methodology introduced in this paper can be applied to various diseases with other epidemic spreading models.
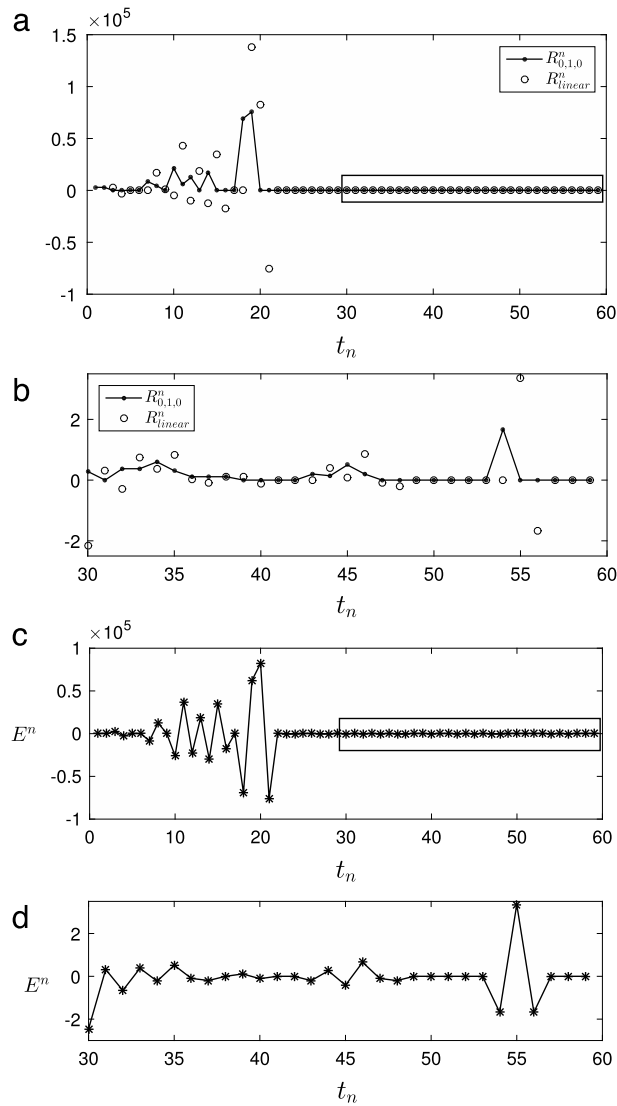
Acknowledgments

Fig. 5. (a) Comparison of the $R_{0,1,0}^n$ with R_{linear}^n which are denoted by the symbols (·) and (◦), respectively. (b) A magnified view of a rectangular box in (a). (c) Error for $R_{0,1,0}^n$ from 20 May 2015 to 17 July 2015. (d) A magnified view of a rectangular box in (c) from 18 June 2015 to 17 July 2015.
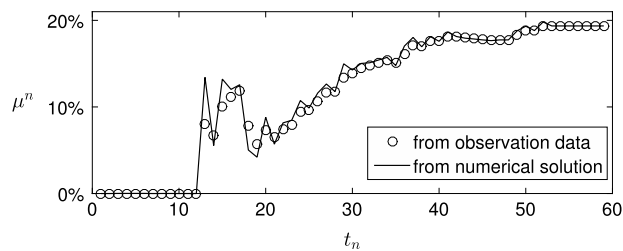

Fig. 6. Fatality percentage rate μ^n by the observation data (circle marker) and the numerical approximations (solid line).

Appendix

In this section, we present the parameter fitting algorithm using the MATLAB function **lsqcurvefit**, which finds the optimal parameters in the least-squares sense. The following command returns the optimal parameters.

$\beta^0 = 1.0e\text{-}6$; $\gamma^0 = 1.0e\text{-}1$;
$[\beta \ \gamma] = \text{lsqcurvefit}(@\text{SIRmodel}, [\beta^0 \ \gamma^0], [n \ n+1], \text{SIRdata}, [0 \ 0], [1 \ 1])$;

In this routine, β^0 and γ^0 are an initial guess, $[n \ n+1]$ is a time interval, and $\text{SIRdata} = [DS^{n+1} \ DI^{n+1} \ DR^{n+1}]$. Also, $[0 \ 0]$ and $[1 \ 1]$ represent the lower and upper bounds for the fitting parameters β^{n+1} and γ^{n+1}. Here, the function SIRmodel is given as follows.

```
function f = SIRmodel(Parameter, [n  n + 1])
    S(1) = DS^n;  I(1) = DI^n;  R(1) = DR^n;
    b = Parameter(1);  g = Parameter(2);
    m = 100;  Δt = 1/m;
    for i = 1 : m
        S(i + 1) = S(i) − b · S(i) · I(i) · Δt;
        I(i + 1) = I(i) + (b · S(i) · I(i) − g · I(i)) · Δt;
        R(i + 1) = a − S(i + 1) − I(i + 1);
    end
    f = [S(m + 1)  I(m + 1)  R(m + 1)];
```

References

[1] X. Zhang, B. Ge, Q. Wang, J. Jiang, H. You, Y. Chen, Epidemic spreading characteristics and immunity measures based on complex network with contact strength and community structure, Math. Probl. Eng. 2015 (2015) 12. Article ID 316092.
[2] B.J. Cowling, M. Park, V.J. Fang, P. Wu, G.M. Leung, J.T. Wu, Preliminary epidemiologic assessment of MERS-CoV outbreak in South Korea, May–June 2015, Euro Surveill. 20 (2015).
[3] Korea Ministry of Health and Welfare (KMOHW) and Korean Centers for Disease Control and Prevention (KCDC). MERS portal (Daily report), http://www.mers.go.kr.
[4] Korea Ministry of Health and Welfare (KMOHW) and Korean Centers for Disease Control and Prevention (KCDC). Daily report for number of infected, recovered, and dead people. http://www.mw.go.kr/, http://cdc.go.kr/CDC/.
[5] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci. 115 (1927) 700–721.
[6] C. Ji, D. Jiang, N. Shi, Multigroup SIR epidemic model with stochastic perturbation, Physica A 390 (2011) 1747–1762.
[7] M. Tanaka, Y. Sakumoto, M. Aida, K. Kawashima, Study on the growth and decline of SNSs by using the infectious recovery SIR model, in: 2015 10th Asia-Pacific Symposium on Inform. Telecommun. Technol. (APSITT), IEEE, 2015.
[8] M.F.C. Gomes, S. Goncalves, SIR model with general distribution function in the infectious period, Physica A 388 (2009) 3133–3142.
[9] E. Tornatore, S.M. Buccellato, P. Vetro, Stability of a stochastic SIR system, Physica A 354 (2005) 111–126.
[10] C. Ji, D. Jiang, N. Shi, The behavior of an SIR epidemic model with stochastic perturbation, Stoch. Anal. Appl. 30 (2012) 755–773.
[11] B. Shulgin, L. Stone, Z. Agur, Pulse vaccination strategy in the SIR epidemic model, Bull. Math. Biol. 60 (1998) 1123–1148.
[12] L. Stone, B. Shulgin, Z. Agur, Theoretical examination of the pulse vaccination policy in the SIR epidemic model, Math. Comput. Modelling 31 (2000) 207–215.
[13] L. Fu, W. Song, W. Lv, S. Lo, Simulation of emotional contagion using modified SIR model: A cellular automaton approach, Physica A 405 (2014) 380–391.
[14] S. Mandal, R.P. Sarkar, S. Sinha, Mathematical models of malaria-a review, Malar. J. 10 (2011) 1–19.
[15] S. Polwiang, The seasonal reproduction number of dengue fever impacts of climate on transmission, PeerJ PrePrints e1142 2015. https://dx.doi.org/10.7287/peerj.preprints.756v2.
[16] C.J. Duncan, S.R. Duncan, S. Scott, Whooping cough epidemics in London, 1701-1812: infection dynamics, seasonal forcing and the effects of malnutrition, Proc. R. Soc. Lond. B: Biol. Sci. 263 (1996) 445–450.
[17] Y. Wang, J. Cao, A. Alofi, A.M. Abdullah, A. Elaiw, Revisiting node-based SIR models in complex networks with degree correlations, Physica A 437 (2015) 75–88.
[18] V.D. Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission, Math. Biosci. 180 (2002) 29–48.
[19] G. Chowell, F. Abdirizak, S. Lee, J. Lee, E. Jung, H. Nishiura, C. Viboud, Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study, BMC Med. 13 (2015) 210–221.
[20] O. Diekmann, H. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, Wiley, New York, 2002.
[21] J.A.P. Heesterbeek, A brief history of R_0 and a recipe for its calculation, Acta Biotheor. 50 (2002) 189–204.
[22] N. Britton, Essential Mathematical Biology, Springer Science & Business Media, 2012.
[23] G. Chowell, S. Blumberg, L. Simonsen, M.A. Miller, C. Viboud, Synthesizing data and models for the spread of MERS-CoV, 2013: Key role of index cases and hospital transmission, Epidemics 9 (2014) 40–51.