2013 4th International Conference on Intelligent Systems, Modelling and Simulation

Infectious Disease Spread Analysis Using Stochastic Differential Equations for SIR Model

Yoshihiro Maki, Hideo Hirose
Department of Systems Design and Informatics
Kyushu Institute of Technology
Fukuoka, 820-8502 Japan
Email: hirose@ces.kyutech.ac.jp

***Abstract*—Pandemic simulation is considered to be crucial as a scenario simulation and it is performed by many kinds of methods; the classical ordinary differential models (SIR model), agent-based models, internet-based models, and etc are among them. The SIR model is one of the fundamental methods to see the behavior of the pandemic with easy computation. However, there are no stochastic variation in the equations. The stochastic differential equations (SDE) can provide such kind of variations. Although the SDE are applied to many fields such as economics, less attention has been paid to the SIR simulations. In this paper, we propose a SDE version of the SIR simulation model with application to SARS (Severe Acute Respiratory Syndrome) case in 2003 in Hong Kong.**

***Keywords*-SIR; stochastic differential equation; pandemic; SARS;**

I. Introduction

Pandemic simulation is considered to be crucial as a scenario simulation because we have very limited experience of real pandemics such as a newly emerged infectious disease spread or an infectious disease spread by terrorism. Considering the social impact due to this could be immense, to reduce the risk of pandemic, it is strongly recommended that we obtain information on the spread of diseases in as many as situations we can imagine in the real world.

A classical method to perform pandemic simulations is to use the ordinary differential equations, called the SIR or SEIR [11], [4], [10], [15]. The SEIR model, where S, E, I, and R denote *s*usceptible, *e*xposed, *i*nfected and *r*emoved populations respectively, is an extension of the SIR epidemiological model, which computes the number of people infected with a contagious disease in a closed population over time. The SEIR model can quickly deal with simulations of infectious disease spread among homogeneous populations using simple simultaneous ordinary differential equations and a few parameters. However, there are no stochastic variation terms in the equations.

The stochastic differential equations (SDE) [1], [3] play an important role in a variety of fields such as physical, economical, and medical phenomena. Using the SDE, we can consider the confidence intervals to the predicted values. However, less attention has been paid to the SIR simulations with the the SDE [5], [6]. The objective of our study is to obtain the confidence intervals for the predicted values using a real world case. In this paper, we propose a SDE version of the SIR simulation model with application to SARS (Severe Acute Respiratory Syndrome) case in 2003 in Hong Kong.

II. The SIR Stochastic Differential Equation

A. Deterministic SIR Model

The original SIR model, in which a Malthusian growth model has been adjusted to by Kermack and McKendrick, is a well-known model for simulating an epidemic growth using the ordinary differential equations (ODE) (1), where S, I, and R are the susceptible, infectious and removed populations, and the parameters β, and γ are the infection rate, and the removal rate (recovery rate) rate. For each time t the equations are described as follows:

$$\begin{aligned} dS(t)/dt &= -\beta S(t)I(t), \\ dI(t)/dt &= \beta S(t)I(t) - \gamma I(t), \\ dR(t)/dt &= \gamma I(t). \end{aligned} \tag{1}$$

We can obtain the cumulative number of infected persons with this model and some numerical method, e.g., the Runge-Kutta method.

B. The SIR model in SDE

We now consider the stochastic version for the SIR model. First, the Ito SDE is defined as

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{2}$$

We will discuss the SDE based on this formula. Adding diffusion coefficients to the three ODE, several patterns might be considered. However, considering that

$$S(t) + I(t) + R(t) = N, \tag{3}$$

where N is a constant number which denotes the summation of all groups, we can also see that

$$dS(t)/dt + dI(t)/dt + dR(t)/dt = 0.$$

Hence, the SIR SDE's model may have one or two diffusion coefficients. Here, we propose, first, a simple case because the study here is mainly focused on to suggest the SIR model

with SDE method. For $dI(t)$ term, adding the diffusion coefficient, we have

$$dI(t) = \beta\gamma I(t)dt - \gamma I(t)dt + \sigma(t, I(t))dW_t \tag{4}$$

where

$$\sigma(t, I(t)) = \alpha I(t).$$

The α is a constant, and W_t denotes 1-dimensional Wiener process.

C. Estimation of Parameters for the Model

The original deterministic model has two parameters. These parameters can be estimated using difference equation methods.

$$\beta(t_i) = \frac{S(t_i) - S(t_{i+1})}{S(t_i)I(t_i)}, \ \gamma(t_i) = \frac{R(t_{i+1}) - R(t_i)}{I(t_i)}. \tag{5}$$

To estimate the parameters of SDE, a well-known property of the quadratic variation associated with the stochastic process for $I(t)$ as follows:

$$\hat{\alpha}^2 = \frac{\sum(\{I(t_{i+1}) - I(t_i)\}^2}{\sum(t_{i+1} - t_i)I^2(t_i)}. \tag{6}$$

Many researchers use this kind of formula, e.g., [8], [9]

III. Numerical Procedure

The important part of this argument is to give a stochastic variation to the original SIR model. As noted previously, we now can simulate the pandemic growth with random process in the SIR model. We will show next how we proceed the simulation and how we define the confidence intervals for simulated process.

A. Numerical method for SDEs

Many numerical methods for SDEs have been studied: Euler-Maruyama method, Milstein method, Stochastic Heun method, and stochastic Runge-Kutta method [7], [12], [3]. We use Euler-Maruyama method as a simple numerical method.

$$X_{t_{i+1}} = X_{t_i} + b(t, X_{t_i})\Delta t + \sigma(t, X_{t_i})\Delta W_{t_i}$$

In the case of SIR SDE, we obtain

$$\begin{aligned} S(t_{i+1}) &= S(t_i) + \{-\beta S(t_i)I(t_i)\}\Delta t_i, \\ I(t_{i+1}) &= I(t_i) + \{\beta S(t_i)I(t_i) - \gamma I(t_i)\}\Delta t_i \\ &+ \alpha I(t_i)\Delta W_i, \\ R(t_{i+1}) &= N - S(t_{i+1}) - I(t_{i+1}), \end{aligned} \tag{7}$$

or

$$\begin{aligned} S(t_{i+1}) &= N - I(t_{i+1}) - R(t_{i+1}), \\ I(t_{i+1}) &= I(t_i) + \{\beta S(t_i)I(t_i) - \gamma I(t_i)\}\Delta t_i \\ &+ \alpha I(t_i)\Delta W_i, \\ R(t_{i+1}) &= R(t_i) + \{\gamma I(t_i)\}\Delta t_i, \end{aligned} \tag{8}$$

where

$$\Delta W_i = W_{t_{i+1}} - W_{t_i}.$$

Note that the random variables ΔW_i are independent and identically distributed normal random variables with expected value zero and variance Δt_i. The Δt_i is a minute time variation. However, as far as we are concerned with SDE, their stability is a subtle problem as the simulated pandemic growth would be unstable if Δt is not small enough [13].

Here, we apply the method to show an example from real world data in the SIR model. In the case of the logistic model, the confidence intervals has been reported via exact analysis. However, such an exact analysis for the SIR model has not yet provided; see [2]. Therefore, we show the confidence intervals by using the numerical simulation approach. In each time, we obtain the distribution for infected people $R(t)$ via the stochastic processes. Thus, the confidence intervals can be obtained by the simulation. We will discuss how many runs we need to simulate in the following section.

IV. Simulation with Real Data Case: SARS

As a real world example case, we deal with the case of SARS. We need to consider the number of simulation runs for obtaining appropriate confidence intervals as mentioned above. Figures 1-3 show us the difference between the number of iterations; results by using 100, 1000, and 10000 runs are shown. In the figures, the simulated curve and its confidence intervals of R whose significance level is 95% with truncated data are shown; the confidence intervals are illustrated as dashed line; the solid line is the result without stochastic processes. The confidence intervals in SDE are computed by combining the many solutions. The Figure 3 seems smooth enough to use 10000 runs.

Figures 3-5 show differences among the observed values of days 1-40, 1-50, and 1-60. The significance level is 95%; each figure has dots, open dots, and three curves. The dots which are on the left side of the vertical line denotes the data for estimating parameters. The open dots are observed data. Among three curves in the figures, the curve in the middle is drawn by using the SIR model without stochastic processes; the others are those of confident intervals for each time. Their confident intervals are obtained by using 10000 times simulations. The parameters selected are, transition day of 3, $N = 10000$, and $dt = 0.1$. We see from Figures 3-5 that the SIR model gives us the final value around 2300 at the time of day 40. This estimated value and the observed value of 1755 are close to each other. However, the confidence interval show a possibility that the number of infected people would be twice as many as the actually observed number. As time goes on, the highest value in 95% confidence intervals, which we can interpret the possible worst case, is becoming lower.

Some researches has reported about the mean and the variance of simulation for stochastic processes. They assume

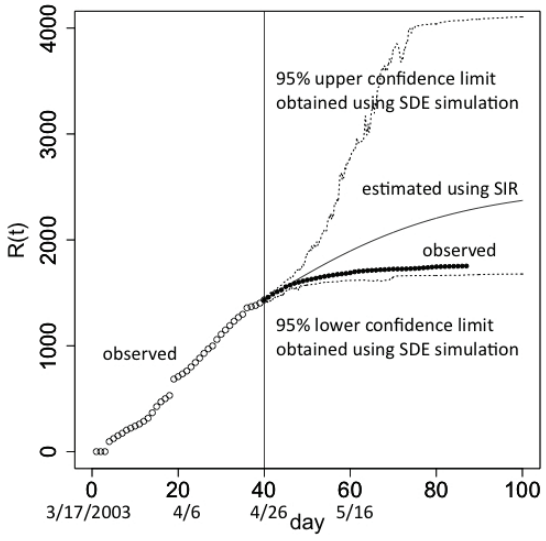

Figure 1. The simulated curve and its confidence intervals of R with 100 runs using observed values of days 1-40.

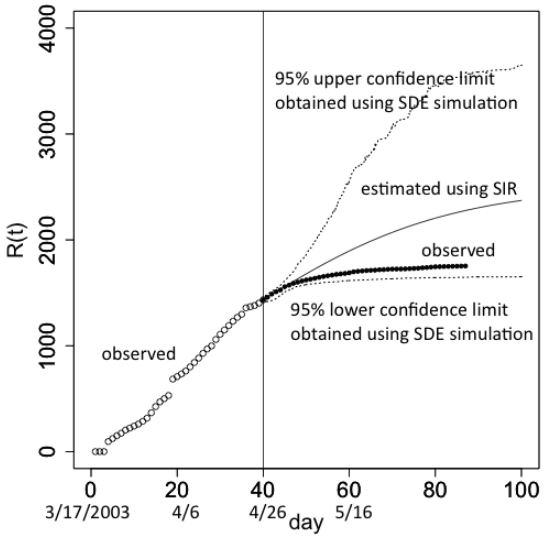

Figure 2. The simulated curve and its confidence intervals of R with 1000 runs using observed values of days 1-40.

one ODE in the SIR model as an SDE, and we cannot established theories to explain how they select functions. We here assume two SDEs in the SIR model. In addition, we provide each diffusive coefficients to combine the functions. For instance, when dS/dt has $\alpha I(t)dW_t$ as diffusive coefficient. Then, we compare some cases:

1) S has diffusive coefficient for S
2) S has diffusive coefficient for I
3) S, I have diffusive coefficient for S, I each.

These simulations are shown in Figures 6-8. In Figure 7, we see that the functions in diffusive coefficient are effective because true values that are not used for estimating parameters are located out of confidence intervals. We can see that the diffusive coefficient $\alpha I dW_t$ in dS/dt is not appropriate. In the case of two SDEs in Figure 8, the simulated epidemic growth curves seem to be unstable due to the width of the confidence intervals.

V. Conclusion

Pandemic simulation is considered to be crucial as a scenario simulation, and its simulation is performed by many kinds of methods. The SIR model is one of the fundamental methods to see the behavior of the pandemic with easy computation. However, there are no stochastic variation in the equations. The stochastic differential equations (SDE) can provide such kind of variations. Although the SDE are applied to many fields such as economics, less attention has been paid to the SIR simulations. In this paper, we propose a SDE version of the SIR simulation model with application to SARS (Severe Acute Respiratory Syndrome) case in 2003 in Hong Kong. We pursued here to obtain the

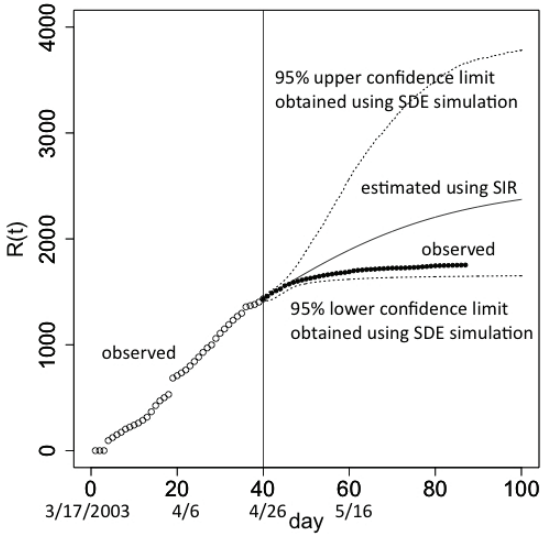

Figure 3. The simulated curve and its confidence intervals of R with 10000 runs using observed values of days 1-40.

confidence intervals for the estimates using many runs of simulations. Using 10000 runs and selecting the diffusion coefficient $\alpha I dW_t$ for dS/dt may be appropriate.

References

[1] B. Øksendal, Stochastic Differential Equations An Introduction with Applications, Springer-Verlag Berlin Heidelberg, 1998.

[2] C.H. Skiadas, Exact Solutions of Stochastic Differential Equations: Gompertz and Generalized Logistic, *Methodology and Computing in Applied Probability*, vol.12, pp.261270, 2009.

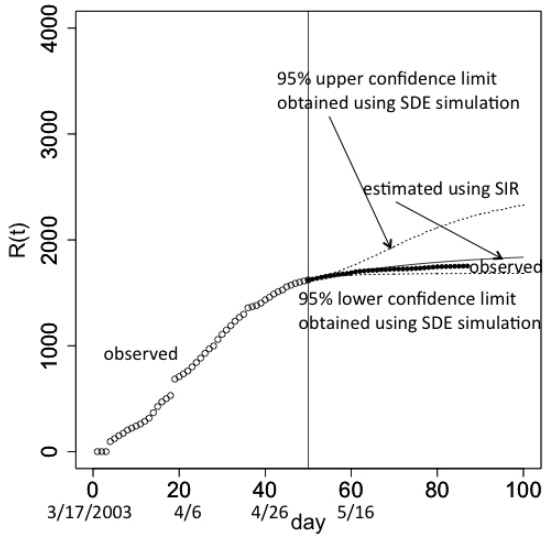

Figure 4. The simulated curve and its confidence intervals of R with 10000 runs using observed values of days 1-50.

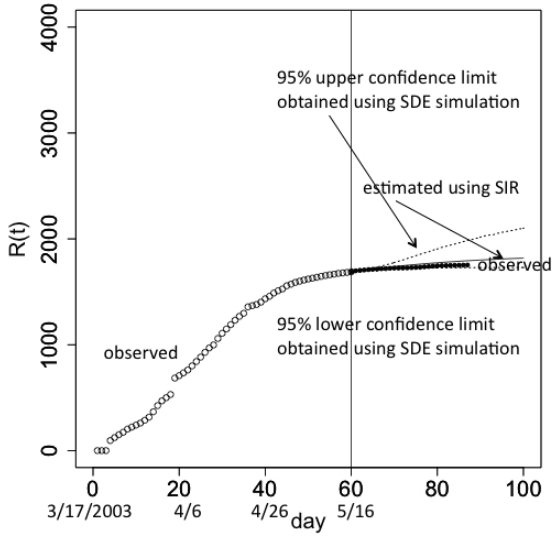

Figure 5. The simulated curve and its confidence intervals of R with **10000 runs** using observed values of days 1-60.

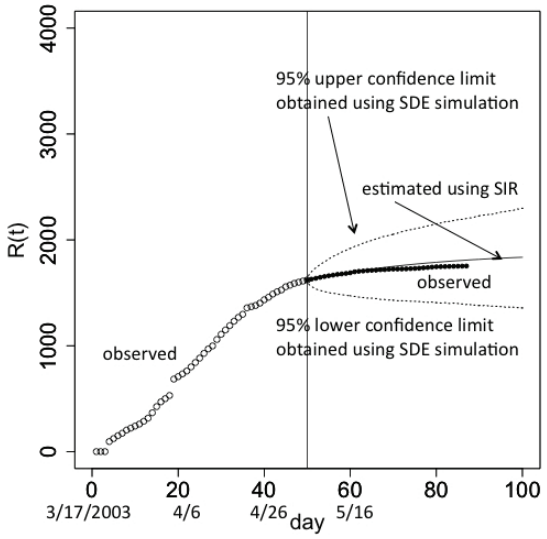

Figure 6. The simulated curve and its confidence intervals of R using observed values of days 1-50, where dS/dt has diffusion coefficient "$\alpha S dW_t$".

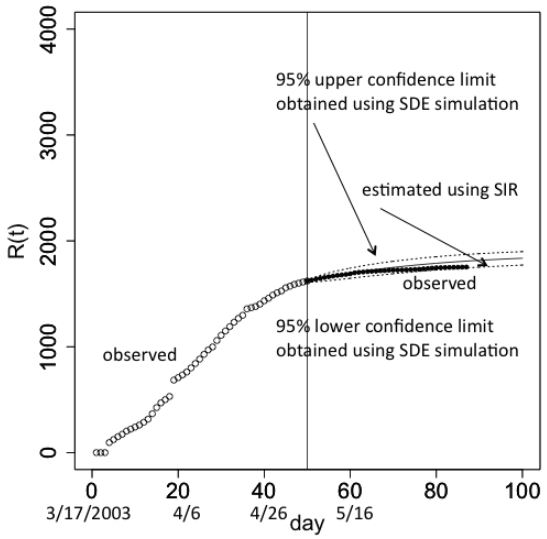

Figure 7. The simulated curve and its confidence intervals of R using observed values of days 1-50, where dS/dt has diffusion coefficient "$\alpha I dW_t$".

[3] D.J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, SIAM REVIEW, Vol. 43, No. 3, pp. 525546, 2001.

[4] F. Brauer, P. van den Driessche and J. Wu (ed.), *Mathematical Epidemiology*, Lecture Notes in Mathematics, Springer, 2008.

[5] J.P. Medlock, The Effect of Stochastic Migration on an SIR Model for the Transmission of HIV, Georgia Institute of Technology, 1999.

[6] J.P. Trapman, On stochastic models for the spread of infections, Vrije Universiteit, 2006.

[7] J. Wilkie, Numerical Methods for Stochastic Differential Equations, Phys. Rev. E, vol.70, issue 1, pp.017701-017704, 2004.

[8] L. Ferrante, S. Bompadre, L. Possati, and L. Leone, *Parameter Estimation in a Gompertzian Stochastic Model for Tumor Growth*, Biometrics vol.56, pp.1076-1081, 2000.

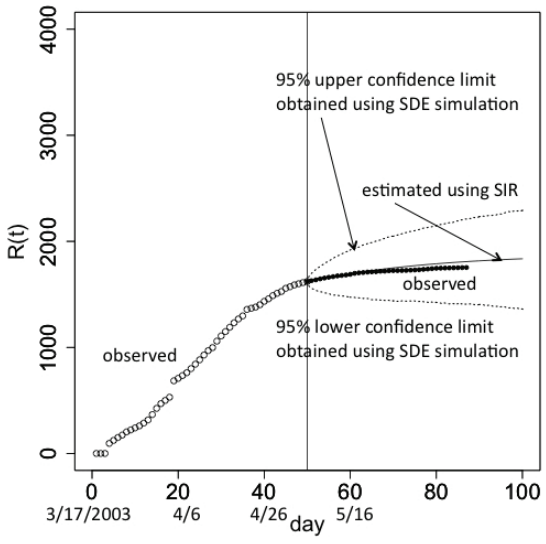

Figure 8. The simulated curve and its confidence intervals of R using observed values of days 1-50, where dS/dt has diffusion coefficient "$\alpha_1 S dW_t$, and dI/dt has diffusion coefficient $\alpha_2 I dW_t$".

[9] M. Hoffmann, Annales de l'Institut Henri Poincare (B) Probability and Statistics, On Estimating the Diffusion Coefficient: Parametric Versus Nonparametric, vol. 37, issue 3, pp. 339372, 2001.

[10] O. Diekmann and J.A.P. Heesterbeek *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, New York: Wiley, 2000.

[11] R. Anderson and R. May, Infectious diseases of humans: Dynamics and control, *Oxford University Press*, 1991.

[12] Y. Saito, The Modified Heun Method for Stochastic Differential Equations, Transaction on JSIAM, vol.21, no.4, pp.323-333, 2011.

[13] Y. Saito and T. Mitsui, Error Analysis of Numerical Scheme for Stochastic Differential Equation, Report of Research Institute for Mathematical Sciences of Kyoto University, vol.850, pp.124-138, 1993.

[14] Y. Toyosaka, H. Hirose, The consistency of the pandemic simulations between the SEIR model and the MAS model, IEICE Transactions on Fundamentals, Vol.E92-A, No.7, pp.1558-1562 (2009.7)

[15] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity, *Proceedings of the Royal Society*, vol. 141A, pp. 94-122. 1933.