Estimation of Macroscopic Parameter in Agent-based Pandemic Simulation

Masaya M. Saito*, Seiya Imoto¹, Rui Yamaguchi†, Satoru Miyano¹ and Tomoyuki Higuchi*
* The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-0014, Japan
† Human Genome Center, Institute of Medical Science, University of Tokyo,
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Abstract—Simulations of epidemic spread is an appealing approach to design intervention programs against influenza epidemic. Agent-based approach is particularly useful, since it enable us to administrate and evaluate effectiveness of various intervention measures in the simulated city. On the other hand, observation date are obtained as macroscopic information. Hence, we need to extract some macroscopic information, to examine the validity of the simulation result. The reproduction number, which indicates how many new infected persons are produced due to one infected person, is often used in epidemiology. This quantity is directly related to coefficients in differential equations in macro simulations, whereas it is non-trivial in agent-based simulations. In this paper, we demonstrate the estimation of the reproduction number from an agent-based simulation result, by assimilating the number of infected persons yielded by an agent-based simulator to the SEIR model, which is a representative macro simulation model.

Index Terms—Agent-based simulation, influenza pandemic, parameter estimation

I. INTRODUCTION

Highly pathogenic avian influenza A(H5N1) viruses may cause a serious pandemic, if the viruses acquire the capability of human-to-human transmission. Our interest is to develop simulators that support planning of interventions against influenza epidemics in some metropolitan area, when such a novel influenza virus is introduced. For example, simulation results for different effectiveness of vaccine and different targets of vaccination are useful information, since the effectiveness of vaccine against a novel virus is uncertain (a general estimation is discussed in the report [1]) and the amount of vaccine that can be prepared is limited.

In pandemic simulations, there are two different ways of modelling: one is based on an ordinary differential equations which describe evolution of the numbers of persons in respective health states, namely certain macroscopic quantities, and the other is the so-called agent-based simulations [2], [3], [4], which simulate activities of individual persons and traffic systems in a city or a regions in concern. Benefits of macro simulations are that computation tasks are small and described variables and parameters can be easily compared with the number of infected persons obtained via confirmatory diagnosis or sentinel observations. For example, the situation of concerned epidemic spread is often described by a constant called basic reproduction number[citation needed]. In a macro simulation model, this constant is directly defined in a terms of parameters in differential equations. However, it is difficult to include the effect of some local events into a macro simulations, such as school closures, vaccinations to some selected populations. Contrastively, agent-based approach enables us to model these local events into the simulation, but a certain effort is required to obtain information which can be compared to the reality, from the simulation result.

In this paper, we estimate the reproduction number as a demonstration of the estimation of macroscopic parameters inherent to the result of agent-based simulation. In order to measure the effectiveness of candidate interventions against epidemic through simulations, we first develop a configuration of simulation which occurs a feasible evolution of epidemic spread. Since epidemiological constraints are often described in terms of the (basic) reproduction number, it is important to estimate this value of the simulation. We propose to use data assimilation technique to the estimation, by regarding infected numbers yielded by an agent-based simulation as a “observation time course”, and employ a set of differential equations, called the SEIR model as a system model. Although our main interest is to use our method to analyse simulation results, it is also useful to connect two kinds of simulations. MADE [5] reduces computational complexity by carrying out the simulation in agent-based approach only in the early stage of epidemic and continuing the simulation in the SEIR model, where parameters of the SEIR model are obtained by simply calculating the difference of the numbers of population in an agent-based model.

The rest of this paper is organised as follows. The SEIR model, which is used as a system model, and our agent-based simulator, which is used to yield virtual observation time course, are introduced in Sections II and III, respectively. The formulation of the state space model and particle smoother algorithm are introduced in Section IV-A and how the SEIR model and the virtual observation time course are adopted to the state space model is explained in Section IV-B. A demonstration of the estimation of the reproduction number is carried out in Section V, and the conclusion is addressed in Section IV.

II. MACRO SIMULATION MODEL

In this section, we briefly introduce the SEIR model [6]. This model describes evolution of the numbers of populations in four different health states: susceptible, exposed, infected, and removed, respectively. Let S, E, I, and R be the numbers
of these populations. Then, the SEIR model is given by the ordinary differential equations:

\[
\begin{align*}
\dot{S} &= -\beta(S/N)I, \\
\dot{E} &= \beta(S/N)I - \alpha E, \\
\dot{I} &= \alpha E - \gamma I, \\
\dot{R} &= \gamma I,
\end{align*}
\]

where \( N \equiv S(0) + E(0) + I(0) + R(0) \) is the conserved quantity (the number of total population), \( \alpha^{-1} \) is the latent period, and \( \gamma^{-1} \) is the infectious period. We assign typical values [7] for these periods \( \alpha^{-1} = 3.5 \) days and \( \gamma^{-1} = 3 \) days. The number of secondary infected persons produced by one primary infected person per unit time is given by \( \beta(S/N) \) and the timescale of infectiousness sustaining is given by \( \gamma^{-1} \). Hence, \( \beta S/(N \gamma) \equiv R(0) \) gives the total number of infected people due to one infected person. In the early stage of an epidemic, \( I, E \approx 0 \) and \( S \approx N \), and hence constant \( R_0 \equiv \beta N/\gamma \) is used to predict whether the number of infected people increase or decrease. Hereinafter, \( R_0 \) is called the basic reproduction number, and \( R \) is simply called the reproduction number. If \( R_0 > 1 \) then the epidemic concerned spreads, and otherwise a sequence of transmission will eventually die out. This is obvious from the sign of \( d(I + E)/dt \) at the early state of an epidemic. For the further interpretation of the SEIR model, see the reference [8].

III. AGENT-BASED SIMULATION MODEL

A. Design of simulator

Our simulator aims particularly at epidemic spread in a metropolitan city. In this environment, the concentration of population in particular places related to human activities (e.g. corporations, trains) affects on the infectious transmission, rather than the concentration in some geometrical regions. Hence, we abstract way from geometrical properties, and model the cities as a set of cells that contain persons in the city. This modellings is contrastive to the case of the entire nation or the region including several nations being targeted. For example, in a pandemic simulation in Southeast Asia [9], the effective radius of circles, inside which infectious disease is transmissible, is one of the most important parameter.

A city dealt in our simulator consists of several local towns, and these towns are connected by trains. Figure 1 is a schematic illustration of data structure and behaviour of the simulator. Places in a town are classified into schools, corporations, supermarkets, parks, and homes. A place contains the numbers of persons in respective health states, on which infectious transmission efficiency in the place depends. A town have the other an array keeping the status of persons living in the town. Each element of the array has a scheduler, the place where the person currently visits, and the health state. We prepare three types of behavioral templates, which mimic activities of employee, students, and housekeepers, and their stochastic variants are given to individual persons as their schedulers.

A pseudo code shown in Fig. 2 describes a procedure of a single step of simulation. Concurrently running threads progress simulation by manipulating object city, which holds the information for the simulated city and allocated in shared memory. We parallelize simulation by doing concurrently operations for a set of persons. Operations for persons consists of those for move and those for progress of disease. Move of a person is determined by only their own schedules, and hence moves of different persons can be done independently each other. We model infectious transmission and recover of disease these by employing the same four health states \( s, e, i, \) and \( r \) as those of the SEIR model and transition between states follows a stochastic process. Let us consider one person who visits at a place where there are \( N \) persons in total and \( I \) persons in state \( i \). Then, the transition probability \( \pi(x \rightarrow x') \) per unit time is given by

\[
\begin{align*}
\pi(s \rightarrow e) &= \beta I/N, \\
\pi(e \rightarrow i) &= \alpha, \\
\pi(i \rightarrow r) &= \gamma,
\end{align*}
\]

where the other transition probabilities to a different state are zero, and parameters \( \alpha, \beta, \) and \( \gamma \) have the same meaning as those of SEIR model, respectively. Each place (object) contains a table of \( \pi(x \rightarrow x') \) (place.key=\( x, \) pr in pseudo code). Move and/or change in health state of a person changes transition probabilities of related places. In our implementation, tables of \( \pi(x \rightarrow x') \) are updated to keep consistency whenever such an event arises (operations marked with a star, in pseudo code). It is possible that change concurrently health states of different persons referring tables of \( \pi(x \rightarrow x') \). However, since persons processed in different thread may move from/to the same place, update of table should be in a critical session. Although this is a potential bottle neck, it is not realised since the number of persons who move or change their health state is much smaller than the total number of persons.
The performance issue in parallelization was discussed in our previous work [10].

The latent period \( \alpha^{-1} \) and the infectious period \( \gamma^{-1} \) are set the same values as those in the SEIR model. The reproduction number (at individual places) is defined in a similar way to the SEIR model. A susceptible person who contact with an infected person for time \( \Delta t \) transit to the exposed state in probability \( \pi(s \to e) \Delta t \), and the infectiousness sustains for time \( \pi(i \to r)^{-1} \). Therefore, the number of infected persons owing to a single infected one is given by

\[
S \cdot \pi(s \to e) \cdot \frac{1}{\pi(i \to r)} = \frac{S \beta}{N} \equiv R \approx \frac{\beta}{\gamma} \equiv R_0,
\]

where the approximation is justified at the early stage of epidemic, with \( S(t) \approx N \). The value of \( R_0 \) ranges typically 1 to 2, and not more than 3 in cases of influenza epidemic [11], [12]. This should be considered a constraint on the value of reproduction number averaged over all places in the city, and there may be a large variety among individual places. We give the mean reproduction number for each kind of places as 0.5 in parks, 1.5 in homes, 0.3 in supermarkets, 1.8 in schools, and there may be a large variety among individual places. We give the mean reproduction number for each kind of places as 0.5 in parks, 1.5 in homes, 0.3 in supermarkets, 1.8 in schools, 2.5 in corporations, and 3.0 in trains, and let places of each kind follow a truncated Gaussian distribution with standard deviation being 10% of the mean. This contrast among kinds reflects the following belief: (i) persons are sparsely distributed in parks, (ii) persons are densely distributed in their home, but infected members may be properly isolated, and (iii) persons have much opportunity to contact each others in schools and corporations.

IV. STATE SPACE MODEL AND PARTICLE SMOOTHER

A. Formalism

The state space model (SSM) [13] is a computational framework to realise data assimilation. Let \( x \) be the vector of simulation variables, \( \theta \) be the vector of parameter variables, and \( y \) be the vector of observed data. Then, the state space model is defined by the stochastic map of \( x \) and \( y \),

\[
\begin{align*}
    x_n &= f(x_{n-1}, \theta) + v_n, \quad v_n \sim p(v_n), \\
    y_n &= p(y_n|x_n, \theta).
\end{align*}
\]

Equations (6) and (7) are called the system model and the observation model, respectively. It should be noted that there are many variants of the SSM and this formulation is a special case of the generalised SSM [14].

We shall use the fixed A-lag Particle smoother algorithm (see [15], [16], [17] for derivation), to estimate the reproduction number from the outcome of the simulator introduced in Section III. This algorithm constructs an ensemble \( \{z_{n-\Lambda}^{(m)}\}_{m=1}^M \) of realised values of \( x_n \) so that they give the Monte Carlo approximation

\[
p(x_{n-\Lambda}|y_1, \ldots, y_n) = \frac{1}{M} \sum_{m=1}^M \delta(x_{n-\Lambda} - z_{n-\Lambda}^{(m)}),
\]

and consists of the following steps:

1) Draw \( M \) particle \( \{z_{n-\Lambda}^{(m)}\}_{m=1}^M \) from the prior distribution \( p(x_0|\theta) \) of initial state variables.

2) Let \( X_n \equiv (x_n, \ldots, x_{n-\Lambda}) \) and \( Z_n \equiv (z_n, \ldots, z_{n-\Lambda}) \). Assume that

\[
\begin{align*}
    \{X_n^{(m)}\}_{m=1}^M &\text{ represents } p(X_n|y_1, \ldots, y_{n-1}), \\
    \{Z_n^{(m)}\}_{m=1}^M &\text{ represents } p(Z_n|y_1, \ldots, y_n).
\end{align*}
\]

Steps below construct \( \{X_n^{(m)}\}_{m=1}^M \) and \( \{Z_n^{(m)}\}_{m=1}^M \) from \( \{X_{n-1}^{(m)}\}_{m=1}^M \) and \( \{Z_{n-1}^{(m)}\}_{m=1}^M \):

a) For \( m = 1, \ldots, M \), draw \( v_{n-1}^{(m)} \sim p(v_{n-1}) \), calculate \( x_n^{(m)} = f(x_{n-1}^{(m)}, v_{n-1}^{(m)}, \theta) \) and let \( X_n^{(m)} = (x_n^{(m)}, \ldots, x_{n-\Lambda}^{(m)}) \).

b) For \( m = 1, \ldots, M \), calculate \( l_n^{(m)} = p(y_n|x_n^{(m)}, \theta) \).

c) For \( m = 1, \ldots, M \), calculate \( w_n^{(m)} = l_n^{(m)} \sum_{m=1}^M l_n^{(m)} \).

d) Resample from \( \{X_n^{(m)}\}_{m=1}^M \) with the probabilities \( \{w_n^{(m)}\}_{m=1}^M \) to obtain \( Z_n \).

e) For \( m = 1, \ldots, M \), apply projection \( Z_n \mapsto z_{n-\Lambda} \) to \( \{Z_n^{(m)}\}_{m=1}^M \). We have \( \{z_{n-\Lambda}^{(m)}\}_{m=1}^M \), which represents smoothed distribution \( p(x_{n-\Lambda}|y_1, \ldots, y_1) \).

B. Application

In this subsection, we describe how our models are adopted to the state space model in Eqs. (6) and (7). In order to distinguish the populations in two models, we denote the agent-based version as \( S_{AG}, E_{AG}, I_{AG} \) and \( R_{AG} \) and the SEIR version as \( S_{DE}, E_{DE}, I_{DE} \), and \( R_{DE} \). Before going to the construction of the system and observation models, we
declare the component of the state and observation vectors:

\[ x_t = (S^{DE}(t), E^{DE}(t), I^{DE}(t), R^{DE}(t), J^{DE}(t), \ln \beta^{DE}). \]

\[ y_t = J^{AG}(t). \]

First, we generate a virtual observation time course from the agent-based simulator. Here, we consider that we have daily reports of new infected persons discovered by confirmatory diagnosis. The simulation counterpart of the number of confirmatory cases \( J^{AG}(t) \) in \( t \)-th day is given by

\[ J^{AG}(t) = \int_{t-1}^{t} \alpha E^{AG}(t') dt'. \]

where \( T \) is the simulation time step. The intended meaning of this formula is that \( E_0 \Delta t \) persons become the infected state, and then feel their illness to go hospital, and they are confirmed as infected patients. Its counterpart in the SEIR model is similarly given by

\[ J^{DE}(t) = \int_{t-1}^{t} \alpha E^{DE}(t') dt'. \]

The simulation model consists of two part, and the first part is the numerical integration from time \( t-1 \) to \( t \) of the SEIR model Eq. 1–4 and the differential form of Eq. (11)

\[ j^{DE} = \alpha E^{DE}, \]

with boundary condition \( j^{DE}((t-1)+) = 0 \). The other part is a discrete update of \( \beta^{DE} \) following to the random walk model,

\[ \ln \beta^D_t = \ln \beta^D_{t-1} + v_{\beta}, \quad v_{\beta} \sim \mathcal{N}(0, \sigma_{\beta}). \]

This stochastic variation of \( \beta \) is responsible for any unexplainable variation in the SEIR model. We shall see that, fitting a simulation path yielded by the SEIR model with varying \( \beta \) to the path of agent-based simulation, one have to allow \( \beta \) a rather change. This is reason why the estimation of the reproduction number is non-trivial.

Considering the confirmatory cases \( J^{AG} \) yielded by agent-based model as observed date follows the Poisson process taking its counterpart in the SEIR model, \( J^{DE} \), as the parameter, we have the observation model,

\[ J^{AG} \sim \text{Poisson}(|J^{DE}). \]

Finally, the estimation of \( R(t) \) is given by the ensemble mean of \( \{z^{m}_t\} \) for the corresponding components,

\[ \hat{R}(t) = \frac{1}{M} \sum_{m=1}^{M} S^{(m)}(t) \exp \left[ \ln \beta^{(m)}_{t} \right] N^{(m)} \gamma^{(m)} \].

V. EXPERIMENT

A. Configuration of model city

We carried out a simulation in a model city, whose statistics are summarised in Table I. For 180 days. Allowable values of the basic reproduction number range from 1 to 3 [11], [12]. Taking these values into account, we give higher values to schools and corporations. If further higher value were given, differences among areas in the pattern of epidemic spread disappear. A preceding study [18] conclude that the peak of epidemic comes after about seven weeks from the beginning of the epidemic. We set parameters so that this time scale is reproduced. The progress of epidemic is affected not only by the latent period \( \gamma^{-1} \) but also by the number of places in towns. In a town with many places, for example, with many corporations, the introduction of the primary infected persons are delayed for some corporations. Our experiment in a previous study [10] indicated that the number of corporations or schools should be an order of \( 10^5 \). However, we have the following problem that, according to the statistics of a real city [19], this scale is adequate for schools, whereas it is too little for corporations (an order of \( 10^2 \)).

Under the above-mentioned condition, we have carried out a simulation for 180 days when 15 initial infected persons are introduced in Area A. This simulation result will be used in a demonstration of the estimation of the reproduction number. The evolution of the number of exposed persons are shown in Fig. 3, for each Area, and for the entire city (scaled by half). The evolution moderately diverges among areas. For example, the peak is located at 50 days in Area A, whereas it spans from 30 days to 70 days.

B. Configuration of particle smoother

The number of particles \( M \) is about 3,130,720 (= 8,192 \times 160 \approx 10^5). The standard deviation \( \sigma_{\beta} \) for random walking of \( \beta^{DE} \) is chosen to be \( \sigma_{\beta} = 0.1 \). The value of the negative log-likelihood value is monotonic to \( \sigma_{\beta} \), but saturates at a certain value of \( \sigma_{\beta} \) (Table II). The chosen value is one at the saturated region. In other words, we chose the system noise so that the time course yielded by the system SEIR model overfits to observation time course. The monotonicity of the log-likelihood against \( \sigma_{\beta} \) indicates that the SEIR model does not have an expressibility of epidemic spread among persons split into a set of local spaces. This matter is discussed in our previous work [20], in which we discussed the expressibility of SEIR-like models against real influenza pandemic observation time courses in Japan. The prior distribution \( p(x_0|\theta) \) is given by a product of independent truncated Gaussian distributions for respective

<table>
<thead>
<tr>
<th>Number of places:</th>
<th>area</th>
<th>school</th>
<th>corporation</th>
<th>park</th>
<th>population</th>
<th>supermarket</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>571,641</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>176,866</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>138,684</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>314,861</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>20</td>
<td>2</td>
<td>44,680</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic reproduction numbers:</th>
<th>train</th>
<th>school</th>
<th>corporation</th>
<th>home</th>
<th>park</th>
<th>supermarket</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.05</td>
<td>0.1</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| log-likelihood             | 1.672,160 | 2.564 | 1.857 | 1.726 |

TABLE I

CONFIGURATION OF THE MODEL CITY.

TABLE II

THE LOG-LIKELIHOOD VERSUS \( \sigma_{\beta} \).
components of $x$, and their means and standard deviations are 
$(25, 10)$ for $E^{DE}(0)$, $(25, 10)$ for $I^{DE}(0)$, and $(2, 0, 1.0)$ for $R$, respectively. The other component of $x$ at $t = 0$, $S^{DE}(0) = $ (population of the city), and $R^{DE}(0) = 0$ consists of the parameter vector $\theta$, with dynamical constants $\alpha^{-1} = 3.5$ days and $\gamma^{-1} = 3.0$ days.

C. Demonstration of Estimation

The result of particle smoothing with $\Lambda = 9$ is observed below. Figure 4 shows the original time course of the confirmatory cases $J^{AG}(t)$ in black and its smoothed version $J^{DE}(t)$ in red. As is already mentioned in Section V-B, the system noise is chosen so that $J^{DE}(t)$ overfits to $J^{AG}(t)$. The estimated time course of the reproduction rate $R(t)$ is shown in Fig. 5.

We observed, in Fig. 5, that the $R(t)$ is initially about 5, and continuously decrease to a value less than unity until 75 days. This initial value $R_0$ is larger than allowable basic reproduction numbers. Hence, we have to use smaller basic reproduction numbers for individual places and/or arrange schedules of individual persons so that the opportunity of contact with the other persons is reduced. The time when $R(t)$ reaches unity roughly coincides with the time of the peak of $J^{AG}(t)$ or $E^{AG}(t)$. This behaviour is understandable in the framework of the standard SEIR model. However, $R^{DE}(t)$ tend to increase toward unity, which does not happen in the SEIR model, in which $R(t)$ is a monotonically decreasing function of $t$. The fact that $R(t)$ increase after get below the unity let us consider a scenario that susceptible persons who fortunately does not contact with or infected by other infected persons for 75 days are supplied to places with infected ones,
and the chain of infectious transmission is sustained in a small scale.

VI. CONCLUSION

We have proposed a method which uses particle smoother to estimate the reproduction number, which is a fundamental quantity in infectious disease, from the result of an agent-based simulation. This method applies the data assimilation technique to the SEIR model, as the system model, and a virtual observation time course is obtained from the agent-based simulation result.

In the demonstration, the reproduction number $R(t)$ is successfully estimated by our method. This quantity summarises the interior state of the simulation. As an example, we can find that the transmission efficiency recovers after the peak of epidemic. In actual usage, we can use the estimate $R(t)$ to evaluated the effectiveness of intervention programmes.

REFERENCES