INFORMATION, Vol.16, No.7(B), July 2013

Pandemic Simulations by MADE: A Combination of Multi-agent and Differential Equations, with Novel Influenza A(H1N1) Case

Hideo Hirose

International Information Institute

INFORMATION
Volume 16, Number 7(B), pp.5365-5389
ISSN 1343-4500

Pandemic Simulations by MADE: A Combination of Multi-agent and Differential Equations, with Novel Influenza A(H1N1) Case

Hideo Hirose

Department of Systems Design and Informatics,
Kyushu Institute of Technology,
Fukuoka 820-8502, Japan
hirose@ces.kyutech.ac.jp

Abstract

Two pandemic simulation approaches are known: the multi-agent simulation model and the differential equation model. The multi-agent model can deal with detailed simulations under a variety of initial and boundary conditions with standard social network models; however, the computing cost is high. The differential equation model can quickly deal with simulations for homogeneous populations with simultaneous ordinary differential equations and a few parameters; however, it lacks versatility in its use. We propose a new method named the MADE which is a combination of these two models, such that we use the multi-agent model in the early stage in a simulation to determine the parameters that can be used in the differential equation model, and then use the differential equation model in the subsequent stage. With this method, we may deal with pandemic simulations for real social structures with lower computing costs. Contrary to the statistical inference method which could not predict the final stage unless abundant information is included, the MADE have a possibility to do that only with the earlier stage information. The newly emerged pandemic, the novel influenza A(H1N1) case in 2009, is dealt with.

Key words: MADE, SEIR, MAS, pandemic, novel influenza A(H1N1), truncated model, Runge-Kutta.

1 Introduction

Pandemic simulation is considered to be crucial as a scenario simulation because we have very limited experiences 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 [26], it is strongly recommended to reduce the risk of pandemic that we obtain information on the spread of diseases in as many as situations we can imagine in the real world.

Recently, the *m*ulti-*a*gent *s*imulation method (the MAS, [2], [10], [11], [21], [24]) is frequently used to overcome this difficult problem. Many simulations are performed using powerful computers in which networks of human-to-human are mimicked as accurately as possible. Assuming that a certain number of persons are infected by a disease in an artificial society, we may estimate the disease propagation using this network simulation method. The MAS can deal with detailed simulations under many kinds of initial and boundary conditions. However, the computing cost soars as the population size becomes larger.

Another classical method to perform pandemic simulations is to use the ordinary differential equations, called the SIR or SEIR ([1], [3], [22]). 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, we cannot set many detailed conditions for the model.

To perform many simulations quickly, we aim to develop a new method, the MADE (*m*ulti-*a*gent and *d*ifferential *e*quations, [32]), that combines these two fundamental models. In the early stage of a pandemic simulation, we propose using the MAS model to determine the appropriate parameters for the SEIR model, and then use the SEIR model in the subsequent stages. Parameters for the SEIR model are determined using the results in the MAS model, where the difference equations naturally derived from the SEIR differential equations are used. SEIR solutions are obtained using the Runge-Kutta method. To confirm the validity of this combined method, simulation results using only the MAS model are compared with those using the SEIR model alone, where the SEIR model parameters are obtained using the whole MAS simulation results. We showed in the previous work, [31] and [34], that the results of the MAS model and the SEIR model are consistent. Using this consistency, we will be able to use the combined method in a variety of situations mimicked to the real world with less computational cost. As far as we know, such an approach is not proposed anywhere. Unlike the statistical parameter estimation results with incomplete data, this proposed approach provides another nice property as well as versatile adaptability for many situations. Even though the parameters are obtained in earlier stages of the pandemic by the MAS, the computational

results for the final removed population size using the MADE are close to those using the MAS model alone. On the contrary, the statistical inference method such as the truncated model could not predict the final stage, unless abundant information is provided.

This paper consists of six sections. After the introductory section, Section 2 briefly mentions the MAS model and the SEIR model. The consistency between the two models is given here. Section 3 describes the proposed model, the MADE, and shows computational results using the method in some virtual city. We evaluate the computational accuracy when the MADE is used. The discussions on the aspect for chance of infection, computing cost, etc. are seen in Sections 4. Section 5 is devoted to the prediction for the novel influenza A(H1N1) spread. Section 6 is the concluding section.

2 Multi-agent Model and Differential Equation Model

In this section, we briefly mention a multi-agent simulation model, MAS, and a differential equation model, SEIR, that will be used in the MADE model. Then, we review the consistency between the MAS simulation results and the SEIR computational results.

2.1 Multi-agent Model

We assume that a disease is propagated by some media such as air or airborne droplet from an infectious person to susceptible persons. The larger the probability of disease infection and the greater the chance of contact with people, the bigger the chance of infecting susceptible persons. The disease spread is affected by parameters such as population size, infection probability, and chance of contact, among others.

First, we describe the typical conditions in the MAS. To begin a simulation for a small city that mimics real cities, we assume three agents for simplicity: office workers, housewives, and students; the patterns of human behavior are shown in Figure 1; similar patterns for a smaller population are shown in [34]. Each office worker has a home where a spouse and children may or may not live together. They gather at their homes at night. In the daytime, office workers go to companies, housewives go to supermarkets or stores, and students go to schools everyday, by train, by bicycle, or on foot. The conditions for each human behavior are shown in Table 1, where $U[a;b]$ denotes the uniform distribution from time a to b, and $N[c,d^2]$ denotes the normal distribution with mean c and variance d^2. For example, an office worker leaves home at 7:02 and

goes to his office by car. He arrives at the office at 7:30. After working at the office, he leaves his office at 19:42, and arrives at home at 20:10. Such times are randomly determined according to the rule in Table 1. In his case, he has a chance to catch a disease either in his working place or at home.

Person-to-person infection probability in every minute is defined by β in infectious areas; see Figure 1 and Table 2. Other underlying parameters, such as the latent period, infectious period and so on, are determined by the uniform distributions or constant values (see Table 2), where the latent period is defined by the time from the initial infection to the point where symptoms of the disease typically appear or are diagnosed, and the infectious period is defined by the period of time during which an infected individual becomes infectious to others [1]. In disease propagation, the important conditions of an individual are susceptible, latent, infectious, and removed conditions. In this paper, the so-called incubation period and the latent period are similarly dealt with, and we use the incubation period here.

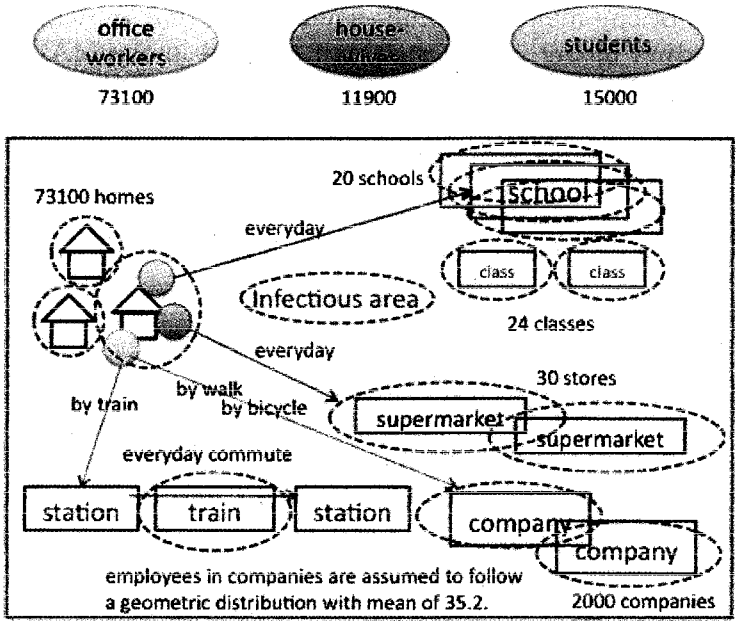

Fig. 1. Patterns of human behavior (case for population size 100,000).

Table 2 also shows the simulation cases. In [34], the populations were set rather small (total 10,000 = 7,000 office workers + 1,500 housewives + 1,500 students) because the focus was on the consistency between the computational results from the MAS model and the SEIR model. Here, a much larger population size is additionally considered, e.g., 100,000 (73,100 office workers + 11,900 housewives + 15,000 students) as shown in Figure 1.

Table 1
Conditions in the MAS model.

population size	10,000	100,000
incubation period	$U[2,7]$ days	$U[2,7]$ days
infectious period	2 days	2 days
office workers		
time of departure (from home)	$U[6:00, 8:20]$	$N[8:00, 2^2h]$
one-way commuting time	$U[1,28]$ min. walk	$U[1,28]$ min. walk
	$U[1,50]$ min. car	$U[1,50]$ min. car
	$U[5,60]$ min. train	$U[5,60]$ min. train
total time spent	$U[400,700]$ min.	-
time of arrival (at home)	$U[12:42, 20:56]$ walk	$N[19:00, 2^2h]$ walk
	$U[12:42, 21:40]$ car	$N[19:00, 2^2h]$ car
	$U[12:50, 22:00]$ train	$N[19:00, 2^2h]$ train
house wives		
time of departure (from home)	$U[10:00, 12:00]$	$N[10:30, 1.5^2h]$
	$U[13:00, 18:00]$	$N[15:30, 2.5^2h]$
one-way commuting time	$U[11,20]$ min.	$U[11,20]$ min.
total time spent	$U[10,30]$ min.	-
time of arrival (at home)	$U[11:12, 20:30]$	$U[11:12, 20:30]$
students		
time of departure (from home)	$U[6:00, 8:20]$	$N[7:15, 1.25^2h]$
one-way commuting time	$U[11,20]$ min.	$U[11,20]$ min.
total time spent	$U[450,480]$	-
time of arrival (at home)	$U[13:32, 16:40]$	$N[17:00, 2^2h]$
trains and stations		
number of cars for a train	10	10
maximum passengers in a car	100	100
interval time between successive trains	8 min.	8 min.
number of stations	10	10

Table 2
Simulation cases in the MAS model.

population size, N	10,000	100,000
number of initial patients, n_0	100, 10, 1	10
infection probability in 1 minute	9.5, 9, 8.5 ,8, 7.5, 7	0.5,0.3
β	$\times 10^{-4}$	$\times 10^{-4}$

2.2 Differential Equation Model

The differential equation models such as the SEIR and the SIR use the ordinary differential equations (1) or (2), where S, E, I, and R are the susceptible, exposed, infectious and removed populations, and the parameters λ, γ, and σ are the infection rate, removal rate (recovery rate), and the transmission rate.

$$\begin{aligned} S'(t) &= -\lambda S(t)I(t), \\ E'(t) &= \lambda S(t)I(t) - \sigma E(t), \\ I'(t) &= \sigma E(t) - \gamma I(t), \\ R'(t) &= \gamma I(t), \end{aligned} \tag{1}$$

$$\begin{aligned} S'(t) &= -\lambda S(t)I(t), \\ I'(t) &= \lambda S(t)I(t) - \gamma I(t), \\ R'(t) &= \gamma I(t). \end{aligned} \tag{2}$$

In the SEIR model, for example, a person could change his or her condition from susceptible to exposed with ratio λ, then to infected after a latent period with a ratio σ, then to removed after an infectious period with a ratio γ. Removed persons will never become susceptible. From equation (1),

$$\sum S'(t) + E'(t) + I'(t) + R'(t) = 0, \tag{3}$$

which means $S(t) + E(t) + I(t) + R(t) = const.$ This is the total population size, and we denote this by N. The Runge-Kutta method is applied to solve (1) or (2).

2.3 Parameters in the Differential Equation Model

Using the MAS model, we obtain the everyday populations for S, E, I, and R. The parameters, λ, γ, and σ for the SEIR model at time t can be approximately

obtained by using the simultaneous difference equations (4).

$$\begin{aligned} \lambda(t) &= \frac{S(t) - S(t+1)}{S(t)I(t)}, \\ \sigma(t) &= \frac{\{E(t) - E(t+1)\} + \{S(t) - S(t+1)\}}{E(t)}, \\ \gamma(t) &= \frac{R(t+1) - R(t)}{I(t)}. \end{aligned} \tag{4}$$

We adopt the values of $\lambda(t)$, $\gamma(t)$, and $\sigma(t)$, assuming that the parameters of the SEIR model are constant and can be estimated as the mean of the corresponding time-dependent quantities. In the SIR model, we can also use equations (4) except for $\sigma(t)$.

2.4 Consistency Between the MAS Results and the SEIR Results

As mentioned in Section 1, it is important to confirm the consistency between the MAS simulation result and the SEIR computational result using the parameters obtained by the MAS. This was shown in the previous work [34] mainly for removed persons, R, where the population size was 10,000. Here, an example of the MAS result and the solution of the SEIR for S, E, I, and R are shown in Figure 2; the population size N is 10,000, infection probability β is 0.01, and the number of initial infected persons n_0 is 10; the SEIR parameters were obtained by using equation (4). From the figure, we can expect the consistency between the MAS results and the SEIR results using the parameters by equation (4).

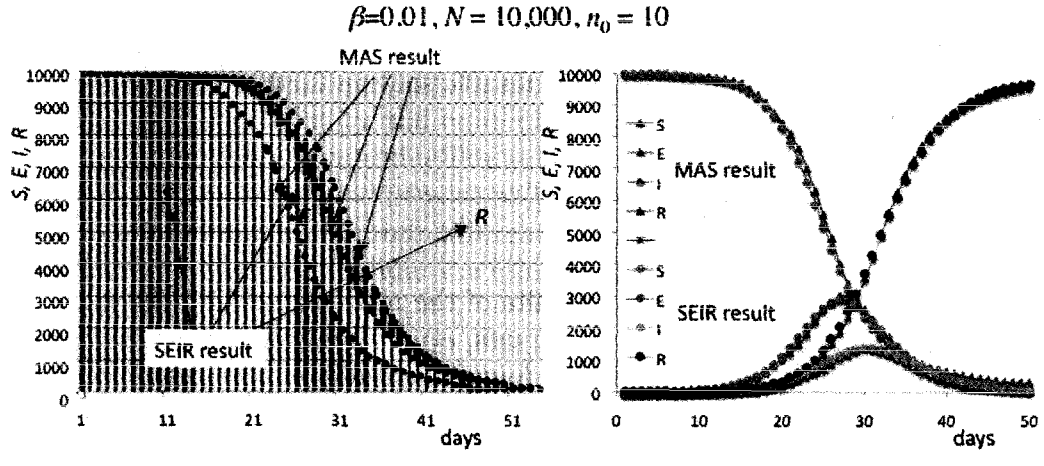

Fig. 2. An example of the MAS result and the solution of SEIR for S, E, I, and R.

To find this consistency to other cases, we introduced an error rate, $e(t)$, or

the ratio of the difference between the SEIR and the MAS to the final value of the MAS. The rate,

$$e(t) = \frac{R(t)_{\mathrm{SEIR}} - R(t)_{\mathrm{MAS}}}{R(t_{\mathrm{final}})_{\mathrm{MAS}}}, \tag{5}$$

was used for accuracy evaluation of the SEIR computational results compared to the results by the MAS simulations. Figure 3 shows the difference of the number of populations between the MAS result and the solution of SEIR for S, E, I, and R for the previous example. Why we adopt the error rate $e(t)$ instead of the commonly used relative error rate is that it is important to grasp the error rate in a global sense rather than in a local sense; the error rate $e(t)$ corresponds to the absolute error.

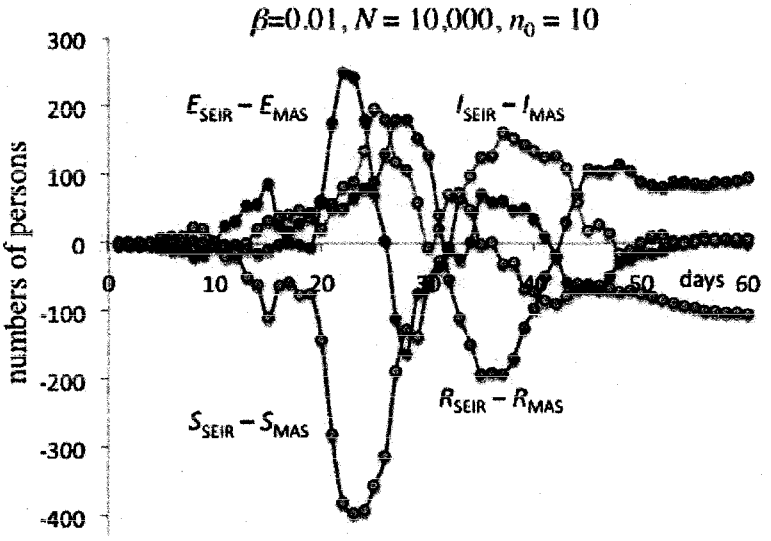

Fig. 3. Difference of S, E, I, and R between the MAS result and the solution of SEIR.

To understand how accurate the SEIR results are, we further used the the maximum error rate, e_∞,

$$e_\infty = \sup_t |e(t)|, \tag{6}$$

as an error norm. According to the results of the mean values and the standard deviations for the maximum error rates, e_∞, by ten simulations to each simulation case (see Table 2), they were found to be very small; in each case, the mean values are less than 0.04 and the standard deviations are less than 0.016 when N is 10,000 [34]. Such a tendency also holds in larger population size simulations.

This indicates the consistency of the computational results between the MAS

and the SEIR. The SEIR can be used as a substitute for the MAS when parameter information are obtained using the MAS. Consequently, it is possible to consider a method that combines these two models.

3 Computation by the MADE

In this section, we first describe the computational method for the MADE, then show the computational results. As explained earlier, the MADE is a combination of the multi-agent (MA) model and the differential equation (DE) model, such that we use the MA model in the early stage in a simulation to determine the parameters that can be used in the DE model, and then use the DE model in the subsequent stage.

3.1 Method of Computation

The MADE is a method to combine the two methods such that we use the MAS in the early stage in a pandemic simulation to determine the appropriate parameters that can be used in the SEIR, and we use the SEIR/SIR in the subsequent stage with parameters obtained in the early stage. The averaging procedure in the SEIR/SIR parameter determination is performed exactly in the same manner as described in 2.3. We use the mean values, $\bar{\lambda}(t)$, $\bar{\sigma}(t)$, and $\bar{\gamma}(t)$ in the MAS part, which can be used as the parameters for the subsequent SEIR/SIR differential equations. For the subsequent stage, the initial values in the SEIR/SIR model for S, E, I, and R are given by the corresponding MAS values on the connecting day.

To perform the pandemic simulations efficiently and effectively, it is important to know when the appropriate connecting time is. Here, we define the connecting time ratio τ by the ratio of the connecting day divided by the cessation time of the pandemic. The cessation time means either 1) the day that the number of removed persons will no more increase, or 2) the day that the number of removed persons will reach the 99% of the converged number of removed persons; we consider the case 2) for R having very long tail.

3.2 Computational Results and Accuracy of the MADE

We investigated the cases of population size 10,000 and 100,000. We use four cases for τ: $\tau = 1/4$, $\tau = 1/3$, $\tau = 1/2$, and $\tau = 2/3$. Figure 4 shows typical cases for the number of removed persons using the MADE; the conditions

are: $N = 10,000$, $n_0 = 100$, $\beta = 9.5 \times 10^{-4}$, and $\tau = 1/4$, $\tau = 1/3$, $\tau = 1/2$, $\tau = 2/3$. We can see that the difference between the results from the MAS and the MADE is very small in this case; see Figure 5 which shows the differences of the number of persons between the two models. In this case, the values of the maximum error rates, e_∞, are 0.0352, 0.0331, 0.0133, and 0.002 for $\tau = 1/4, 1/3, 1/2, 2/3$ cases, respectively. These values are corresponding to the (absolute) peak values in Figure 5. Such small errors are considered to be very useful in efficient and effective pandemic simulation.

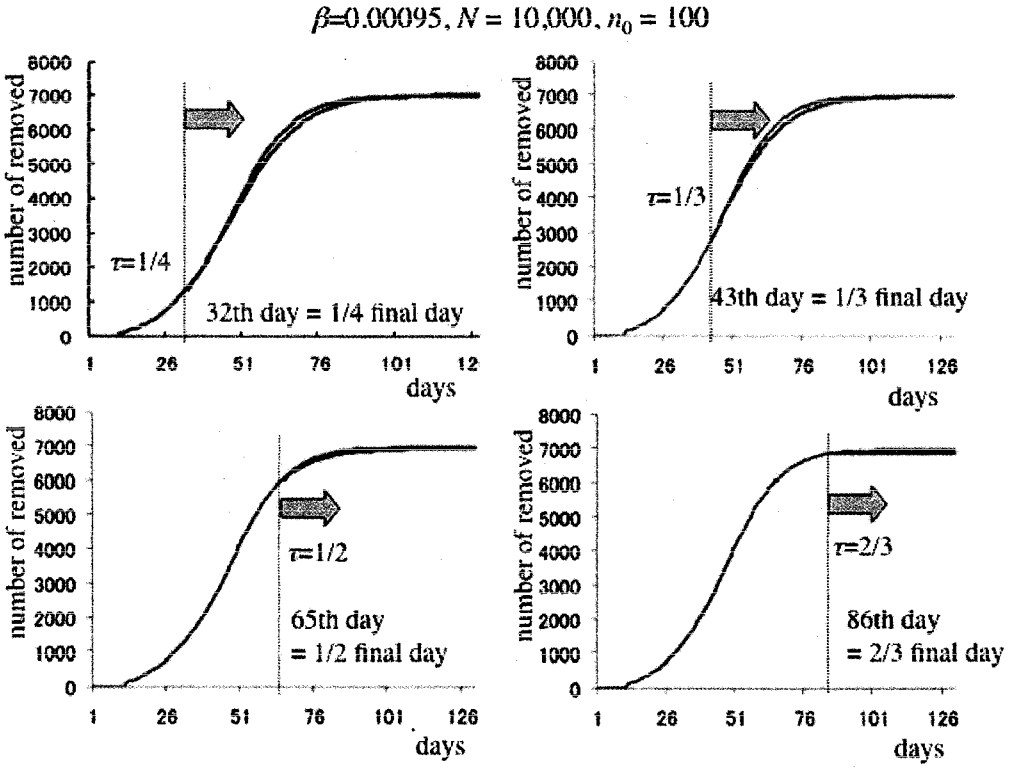

Fig. 4. Number of removed persons using the MADE compared with those using the MAS.

We have checked if this tendency holds in other cases. Table 3 shows the mean values and standard deviations for the maximum error rates e_∞, where $n_0 = 100, 10, 1$, $\beta = 9.5\times10^{-4}, 9\times10^{-4}, 8.5\times10^{-4}, 8\times10^{-4}, 7.5\times10^{-4}, 7\times10^{-4}$, for $N = 10,000$ cases, and $n_0 = 10$, $\beta = 0.5\times10^{-4}, 0.3\times10^{-4}$, for $N = 100,000$ cases, with $\tau = 1/3$, $1/2$, $2/3$; they were computed from ten-time simulations to each case. We can see that the computational results by the MADE do not largely differ from the MAS results; refer to the statistical results shown later, which is different from this property.

We define r as the ratio of the number of difference between the removed persons in the final steady stage $R(t_{\text{final}})_{\text{MADE}}$ and $R(t_{\text{final}})_{\text{MAS}}$ to the number of removed persons in the final steady stage $R(t_{\text{final}})_{\text{MAS}}$,

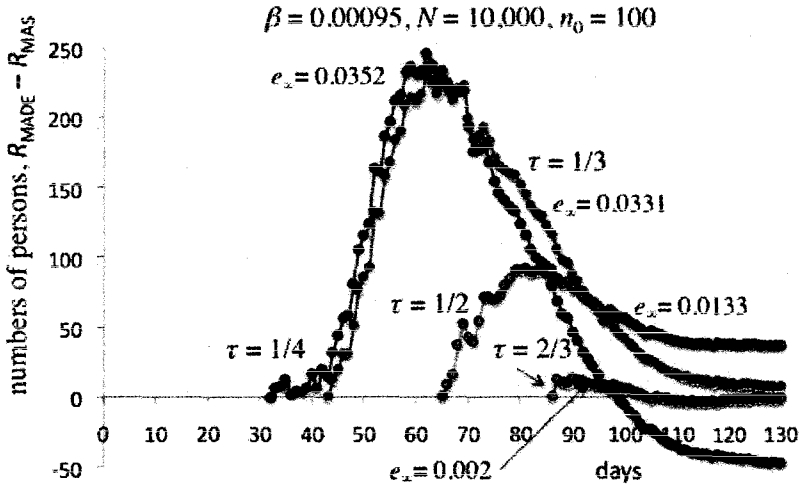

Fig. 5. Difference of the number of persons between the MADE and the MAS.

$$r = \frac{R(t_{\text{final}})_{\text{MADE}} - R(t_{\text{final}})_{\text{MAS}}}{R(t_{\text{final}})_{\text{MAS}}}. \tag{7}$$

This is to see the prediction accuracy for the final stage. Table 4 shows the results for these values. Even if the connecting time ratio τ is 1/3, the values of r are small when n_0 is 100 which corresponds to 1% of the total population. This also suggests the usefulness of the MADE method.

4 Discussions

We discuss here why we introduce the MADE model and what are the pros and cons. If we focus on the prediction to final stages, the MADE provides reasonable answers. Here, we discuss briefly four points: aspect for chance of infection, computational cost, comparison to the truncated model approach, and toward real pandemic cases.

4.1 Aspect for Chance of Infection

The MAS provides the abundant information about how the infectious disease propagated in the network even by a single simulation case, contrary to the SEIR homogeneous simulations. To understand the disease spread phenomena in detail, the MAS approach is an appropriate tool. However, the differential equation model works to predict the final pandemic stages roughly, i.e., to know the magnitude of pandemics. First, we discuss this in a aspect of the chance of infection from simple situations.

Table 3
Maximum error rate, e_∞, by the MADE.

		N	10,000					N	100,000
τ	n_0	β						β	
		9.5	9	8.5	8	7.5	7	0.5	0.3
				$\times 10^{-4}$				$\times 10^{-4}$	
1/3	100	0.0347	0.0440	0.0477	0.0438	0.0480	0.0572		
		0.0128	0.0132	0.0300	0.0176	0.0245	0.0423		
	10	0.0724	0.0842	0.0910	0.0797	0.1145	0.2031	0.1256	0.0508
		0.0315	0.0451	0.0567	0.0437	0.0584	0.0895	0.0543	0.0356
	1	0.1420	0.1823	0.1603	0.1908	0.2247	0.2977		
		0.0824	0.1128	0.0972	0.0946	0.1199	0.1661		
1/2	100	0.0115	0.0131	0.0144	0.0153	0.0117	0.0279		
		0.0084	0.0039	0.0085	0.0072	0.0062	0.0166		
	10	0.0281	0.0315	0.0298	0.0203	0.0336	0.0748	0.0425	0.0163
		0.0144	0.0157	0.0245	0.0117	0.0275	0.0509	0.0272	0.0147
	1	0.0579	0.0647	0.0518	0.0698	0.1065	0.1386		
		0.0385	0.0324	0.0273	0.0291	0.0546	0.0922		
2/3	100	0.0027	0.0033	0.0034	0.0037	0.0044	0.0063		
		0.0013	0.0017	0.0016	0.0023	0.0040	0.0034		
	10	0.0049	0.0041	0.0075	0.0061	0.0139	0.0400	0.0337	0.0148
		0.0033	0.0015	0.0055	0.0038	0.0114	0.0415	0.0193	0.0091
	1	0.0122	0.0201	0.0205	0.0156	0.0341	0.0463		
		0.0112	0.0120	0.0222	0.0076	0.0256	0.0351		

We investigated where the infection occurred in our simulation study. Table 5 shows an example for the numbers of persons and places of infection when $(N, \beta, n_0) = (10,000, 9.5 \times 10^{-4}, 10), (100,000, 0.5 \times 10^{-4}, 10)$. This indicates that people will be infected at the places where they spend time most with infected persons. Thus, the shutdown of human-to-human network will work to prevent from the spread of diseases. This characteristic is also explained in the A(H1N1) influenza case [33]. Why the infected numbers in trains are different from each other between $N = 10,000$ and $N = 100,000$ is due to the unchanged conditions for trains; there may be less crowded cars for $N = 10,000$ and jam-packed cars for $N = 100,000$.

Table 4
Ratio r, number of difference between the removed persons in the final steady stage by the MADE and that by the MAS to the number of removed persons in the final steady stage by the MAS.

		N	10,000					N	100,000
τ	n_0	β						β	
		9.5	9	8.5	8	7.5	7	0.5	0.3
				$\times 10^{-4}$				$\times 10^{-4}$	
1/3	100	0.0264	0.0385	0.0473	0.0376	0.0343	0.0514		
		0.0191	0.0194	0.0290	0.0244	0.0351	0.0469		
	10	0.0666	0.0840	0.0818	0.0653	0.0981	0.1734	0.0159	0.0319
		0.0369	0.0450	0.0564	0.0477	0.0762	0.1223	0.0120	0.0198
	1	0.0351	0.0613	0.0813	0.0811	0.1326	0.1463		
		0.0952	0.0981	0.1008	0.1146	0.1327	0.2323		
1/2	100	0.0030	0.0109	0.0106	0.0033	0.0059	0.0156		
		0.0109	0.0045	0.0112	0.0132	0.0097	0.0266		
	10	0.0079	0.0213	0.0264	0.0132	0.0261	0.0821	0.0203	0.0159
		0.0180	0.0239	0.0267	0.0181	0.0358	0.0613	0.0083	0.0148
	1	0.0042	0.0320	0.0259	0.0280	0.0883	0.1252		
		0.0471	0.0414	0.0450	0.0450	0.0751	0.0943		
2/3	100	0.0008	0.0001	−0.0001	0.0027	0.0015	0.0021		
		0.0023	0.0019	0.0030	0.0029	0.0054	0.0061		
	10	0.0003	0.0002	0.0020	0.0002	0.0074	0.0345	0.0267	0.0169
		0.0034	0.0025	0.0084	0.0066	0.0160	0.0455	0.0048	0.0109
	1	−0.0022	0.0024	0.0041	−0.0010	0.0181	0.0123		
		0.0087	0.0127	0.0125	0.0095	0.0349	0.0493		

Figure 6 shows an example case in which numbers of infected workers, students, and housewives are illustrated. Here, $N = 10,000$, $\beta = 9.5 \times 10^{-4}$, and $n_0 = 10$. We can see that the trends for each infected group are not simply represented; multi-modal trends are observed. However, the cumulative number of infected persons in the MAS model can be approximated well to that in the SEIR model according to the maximum error rate investigation [34]; when $N = 10,000$, $\beta = 9.5 \times 10^{-4}$, and $n_0 = 10$, mean and standard deviation for e_∞ are 0.025 and 0.011, respectively.

Table 5
An example for the numbers of persons and places of infection.

$N = 10,000$, $\beta = 9.5 \times 10^{-4}$, $n_0 = 10$

	office workers	students	housewives
company	4512	0	0
classroom	0	1186	0
home	398	264	926
stores	0	0	12
train	0	0	0

$N = 100,000$, $\beta = 0.5 \times 10^{-4}$, $n_0 = 10$

	office workers	students	housewives
company	58774	0	0
classroom	0	12733	0
home	8331	2245	10984
stores	0	0	177
train	148	0	0

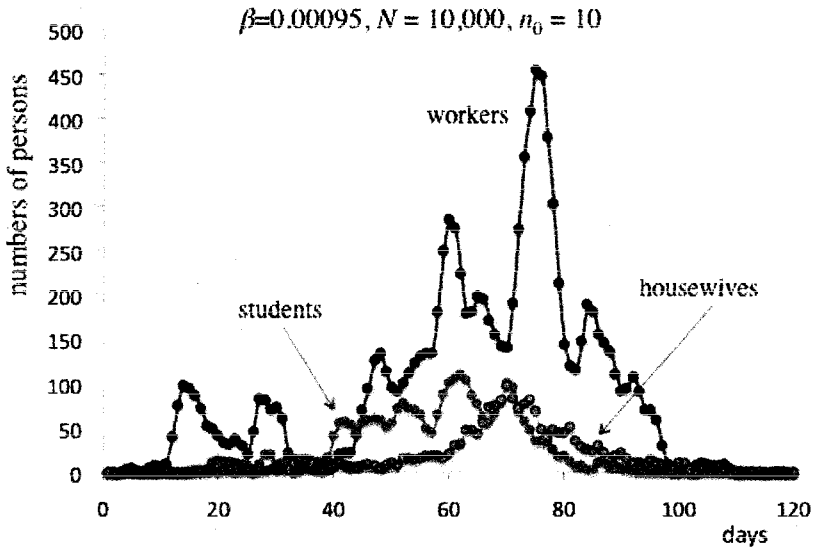

Fig. 6. Numbers of infected workers, students, and housewives.

To understand how infection probability works, we showed, in [34], an example using $\beta = 9.5 \times 10^{-4}$ for $N = 10,000$, where the median time to infection from an infected person to a susceptible person is $(\log 2)/\beta$, and it was ap-

proximately 12 hours. If a susceptible person and an infectious person spend time together for 8 hours, the probability of infection becomes about 0.37. For $N = 100,000$, much smaller values of $\beta = 0.3 \times 10^{-4}, 0.5 \times 10^{-4}$ are used here; this is because the infection chance may become larger among many persons.

4.2 Computing Cost

As introduced in [34], the computing cost in the SEIR model is irrelevant to N, the number of total population, whereas in the MAS model, the computing cost is proportional to N^2 at the worst-case because of the network. In our simulation model, the computational cost could be reduced because there are no infectious propagations between the isolated groups; if the number of groups is G and the size of the groups are the same, then the computing cost should scale as $G \times (N/G)^2 = N^2/G$. The order of the computational complexity is in the range of $O(N)$ to $O(N^2)$. Assuming a worst-case, we have to tackle cases where the population size is about one million, as in real cities like Fukuoka in Japan. As the complexity of the MAS increases, so does the computing cost with N^2 order. The MADE method is, therefore, important in reducing computing cost if the accuracy can be assured. When $N = 10,000$, the reduction rate from the MAS to the SEIR is 3.9×10^{-5} [34]. Therefore, when $N = 1,000,000$, the computing time in the MAS computation becomes from 10^2 to 10^4 times longer, and the reduction rate is about from 4×10^{-7} to 4×10^{-9}. However, computational cost reduction is not the primary objective of the MADE. The important point is that we have shown that even a simple approach described by a few classical differential equations is still attractive for urgent pandemic predictions and can be trusted.

4.3 Comparison to the Truncated Model Results

In the statistical analysis of incomplete data such as SARS, there is a difficulty, to some extent, in estimating the underlying parameters when the censoring time is earlier. According to [14] and [15], when the censoring time is before the half time of the final steady stage, the maximum likelihood estimate of the parameter for the case fatality ratio is not stably obtained. From Fig. 2 in [14], it can be seen that the parameters are not steadily obtained before May 5, 2003; this day corresponds to the point in the logistic curve in Fig. 1 in [14] for the cured and dead persons where the peak time in the density function has passed, and in the curve for the infected persons where they are almost at the final steady stage. This indicates that the statistical estimation procedure is very sensitive to the early censoring time in the truncated model or trunsored model; see also [9], [23], [27] for estimability of the maximum

likelihood parameters.

On the contrary, the MADE can provide rather accurate values for the number of removed persons even though τ, which corresponds to the censoring time, is 1/3. We have yet to ascertain why this holds. One reason may be that the number of unknown parameters are only three compared to the number of equations. The total number of parameters in the statistical model is six in [14] with three kinds of observations: the infected, the cured, and the dead persons. Another reason may be that we know the starting time for the MADE model, which is not the case in SARS. Finally, we guess that the MADE model may be able to inherit the behavioral structure for the epidemiological disease spread at the connecting point with well described differential equations; although solutions are sensitive to parameters, the equations provide deterministic solutions.

Figure 7 shows the prediction results using the maximum likelihood parameter estimation applied to the same example in Figure 4. The dotted curve means the removed numbers obtained by the MAS. The solid curve indicates the predicted curve. We assumed here the three-prameter logistic distribution function as the underlying distribution such that

$$F(t; s, m, c) = [1 + \exp(-\frac{t-m}{s})]^{-c}, \tag{8}$$

where s, m, and c denote the scale, location, and shape parameters. This model is often used in many biological applications such as the growth curve. The literature, [14] and [15], also used this model. Using the truncated model where censoring time is adapted to the connecting time τ, the fitted parameters are, e.g., $\hat{s} = 26.3$, $\hat{m} = -174$, and $\hat{c} = 5357$ for $\tau = 1/3$, and $\hat{s} = 10.3$, $\hat{m} = 43.9$, and $\hat{c} = 1.20$ for $\tau = 1$. Here, the log-likelihood function is denoted by

$$\log L = \log \sum_i \frac{F(t_{i+1}) - F(t_i)}{F(T)}, \tag{9}$$

where T is the censoring time, and t_i denotes the ith day from the beginning. As easily seen by these estimated parameters, we know that the statistical inference fails to obtain the appropriate parameter values. Figure 7 indicates this tendency. In this case, the values of the maximum error rates, e_∞, are 0.231, 0.610, 0.190, and 0.013 for $\tau = 1/4, 1/3, 1/2, 2/3$ cases, respectively; these values are considerably larger than those obtained by the MADE. It seems that the MADE approach is superior to the statistical method using the simulation condition shown in Table 1. We know that this comparison is unfair in a rigorous sense because the original data are derived from the networked human behavioral model and not from the statistical growth model, even if these two are sometimes explained with each other, see e.g., [29]. However,

we have experienced that this example is not a special case but the general case; the A(H1N1) influenza case as seen below (see also [35]) and the foot-mouth-disease (FMD) case emerged in 2009 in Japan have a similar property [16].

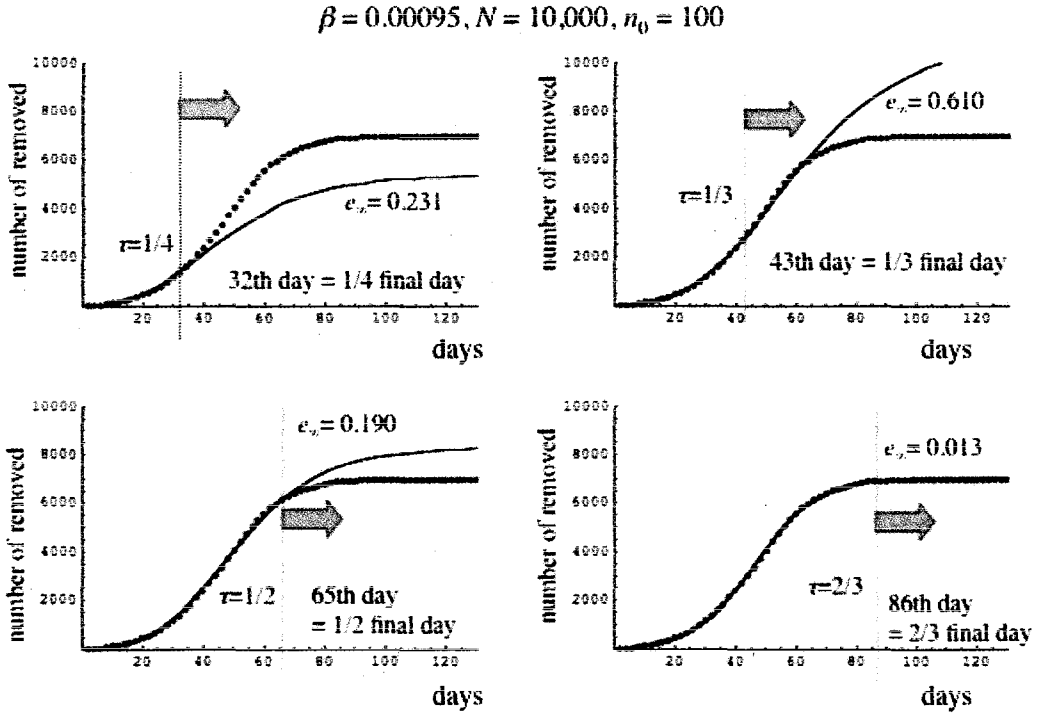

Fig. 7. Number of removed persons using the truncated model compared with those using the MAS.

4.4 From Scenario Simulation to Real Pandemic Prediction

So far, we have explained that the MADE is efficient and effective for final stage predictions of pandemics in the case of scenario simulations. However, when a real pandemic occurred, we cannot easily prepare network models in the real world. The multi-agent simulation is too complex to do that. Traditionally, the SEIR/SIR models have often been used in real pandemic cases. However, the SEIR/SIR computation still relies on the strong condition of homogeneity; that is, it is still theoretical and deductive. As far as we know, this is the first time that the differential equation approach is truly effective even in complex network models which mimic the real human behaviors. Therefore, we may trust the SEIR/SIR computational results in the real world pandemics. If we use the observed data from real pandemics as we have used the MAS data in earlier stages in MADE simulations, we may trust the final stage results using the simple SEIR/SIR models because we may interpret the observed data, S, E, I, and R, as integrated data from a complex network

system. Thus, the conventional use of the SEIR/SIR simulations now have a significant meaning as illustrated in Figure 8. Using real data from the observations like an assimilation, the computational results by the SEIR/SIR become realistic. Based on this standpoint, much more complex use of the SEIR compartment models are expected to provide reliable estimates, e.g., [6] and [12].

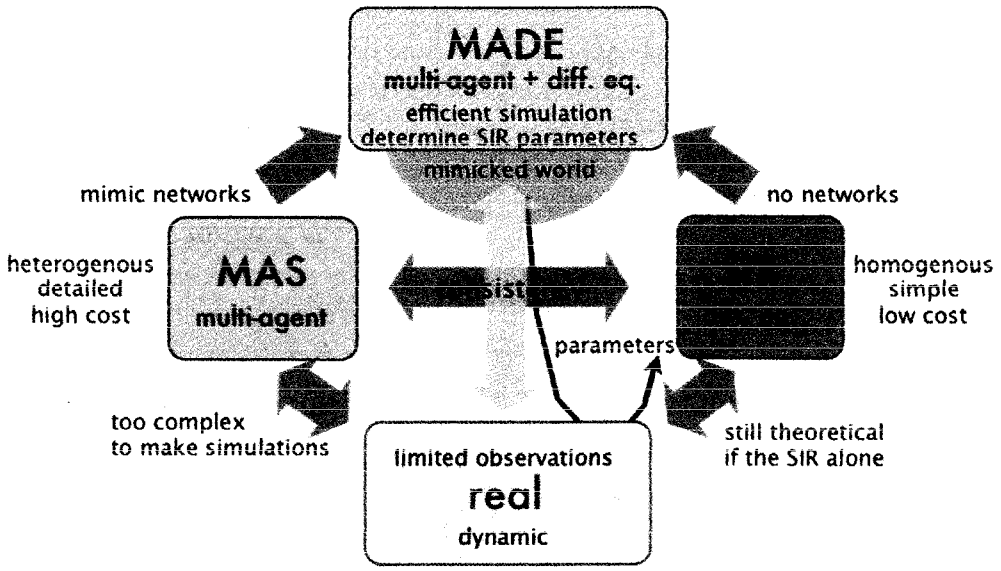

Fig. 8. From scenario simulation to real pandemic prediction.

5 A Case of the Novel Influenza A(H1N1) 2009

In March 2009, WHO alerted the phase of the pandemic of the H5N1 avian influenza: The world is presently in phase 3, a new influenza virus subtype is causing disease in humans, but is not yet spreading efficiently and sustainably among humans. We are on a threat of the influenza pandemic. Since April 2009, the situation has been drastically changed. The novel influenza (so-called swine flue) observed in Mexico, and the disease has spread to all over the world very quickly. On April 27 the phase shifted to 4; on April 29 it raised to 5, and on June 11, it became a pandemic situation, phase 6. The WHO has shown us the number of laboratory-confirmed cases of the pandemic A (H1N1) 2009 from April 24, 2009 continuously. However, after July 6, 2009, the WHO stopped reporting individual laboratory-confirmed cases because of the difficulty in collecting the laboratory-confirmed cases. The influenza, however, spread after all. On August 10 the WHO announced that the A (H1N1) influenza event has moved into the post-pandemic period. Although the disease is not serious comparing to SARS, the number of dead persons due to the novel influenza soars more than 18,449 all over the world by August,

2010 [36], which is higher than the SARS case (774 person died) [37].

Pandemic simulation is a kind of simulation by scenario. We do such a simulation because we rarely meet with a real pandemic. However, we have just encountered such a pandemic in 2009, and learned much from this experience.

1) The outbreak is occurred very suddenly. No one knows what happened for a while. Then, the viruses spread without any protections, resulting in the spread to all over the world in a month. This means that outbreaks occur anywhere anytime, and the disease spread is as quickly as the jet speed from country to country. Animations can be seen via YouTube [38, 39].

2) A kind of panic was seen in many countries in early days both in public and in government. Too strict protections were first planned, which is known to be ineffective later for preventing the disease spread perfectly. However, closing the classes or factories shows somewhat effective for very premature stage.

3) People become insensitive to the new virus when the disease is known not to be serious. There are few people who care about the disease spread. Then, the spread is gradually grown up exponentially, resulting in the pandemic.

These phenomena will be very useful in analyzing the next coming pandemic. For simplicity, we use the SIR model here. The parameters can be obtained by using equation (3), but we only know the number of infected persons. The number of susceptible persons can be found by the population size of the target country. The numbers of infected individuals are reported by the WHO [36] or the CDC [4], but removed populations are not known. Thus, we assume that the infected people will be changed to be removed people in a few days, say three days (the average generation time, i.e. the mean delay between the time of infection of the index case and the time of infection of secondary cases, irrespective of the setting, is 2.5 - 3 days); this is due to the report in [28].

Figure 9 shows the daily estimated parameters for $\lambda(t)$ and $\gamma(t)$ in Japan. Using these values, we can solve the differential equations as shown in Figure 10. The figure shows the computational results by using the SIR as well as the observed infected numbers in Japan. The dots mean the beginning days for SIR model; that is, using the data by that day, the parameters are estimated. Bold curve is the observed infected cumulative numbers. Only the laboratory-confirmed cases officially opened by the WHO by the date of July, 2009 [36] are used for the computation. We see that the results by the SIR agree well with the observed data in a limited duration. The mean values of $\lambda(t)$ and $\gamma(t)$ from the beginning to July, 2009 are $\bar{\lambda} = 2.9 \times 10^{-9}$, $\bar{\gamma} = 0.32$. A final stage situation could be estimated by using the basic reproduction rate $\rho = (\lambda/\gamma) \times S(0)$ [8]. Using $\bar{\lambda}$, $\bar{\gamma}$, and $S(0) = 127,156,000$ (Japanese population), ρ becomes 1.15, and the final number of removed persons in Japan could be 1/4 of the total population by the formula,

$$R(\infty) = S(0)\exp(\frac{1}{\rho} - \rho). \tag{10}$$

This numerical solution told us that the cessation time of the pandemic in Japan could be by March, 2010, if no quarantine, no shutdown of human networks, or no vaccination is operated. In the world case, by a similar treatment, this might be about 1/10 of the total population of the world in two years; this large number of patients and the cessation time estimation are coincident with those in the CDC report [5]. This is also shown in [35].

Now, we know the final stage of this pandemic; Figure 11 shows the comparison of prediction result against the observed one. Although this appears not close to the observed one, the predicted curve can be acceptable, considering the effects of vaccination and network shutdown, even if the connecting point (censoring point) is very early and the number of observed data is small. Lastly, we briefly mention the incompetence of the truncated model. Figure 12 shows the truncated model prediction results; this figure shows that the truncated model estimates the final number of patients to be too moderate.

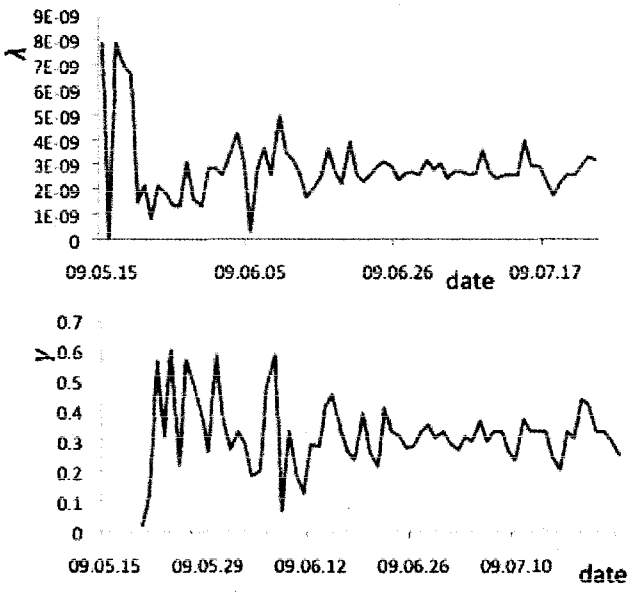

Fig. 9. Daily parameter values of $\lambda(t)$ and $\gamma(t)$ for Japan novel influenza case.

6 Concluding Remarks

Assuming that a certain number of persons are infected by a disease in a typical city and that the disease transmission is described by the MAS model, we can obtain results for the disease spread. On the basis of these results, we can also obtain parameters for the SEIR model using the difference equations.

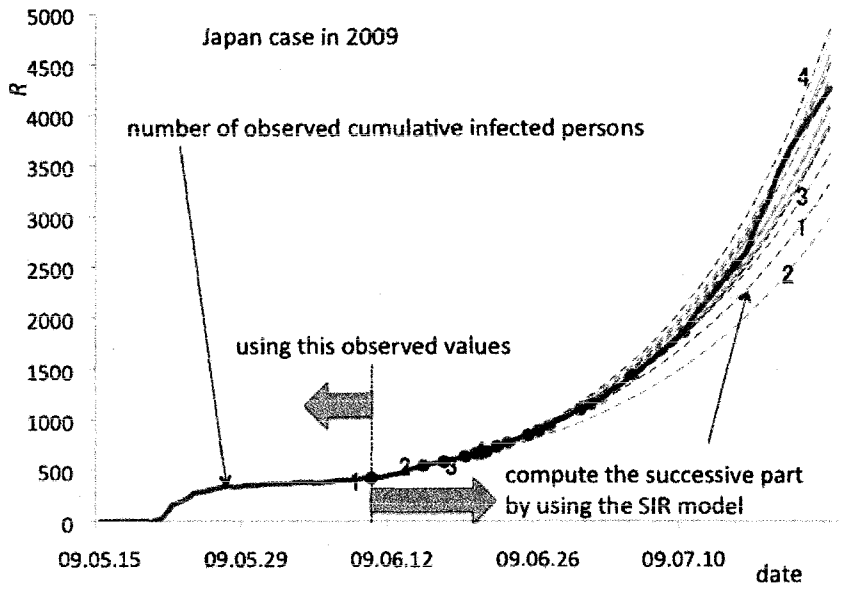

Fig. 10. Number of removed persons using the differential equations for Japan novel influenza case: early stage.

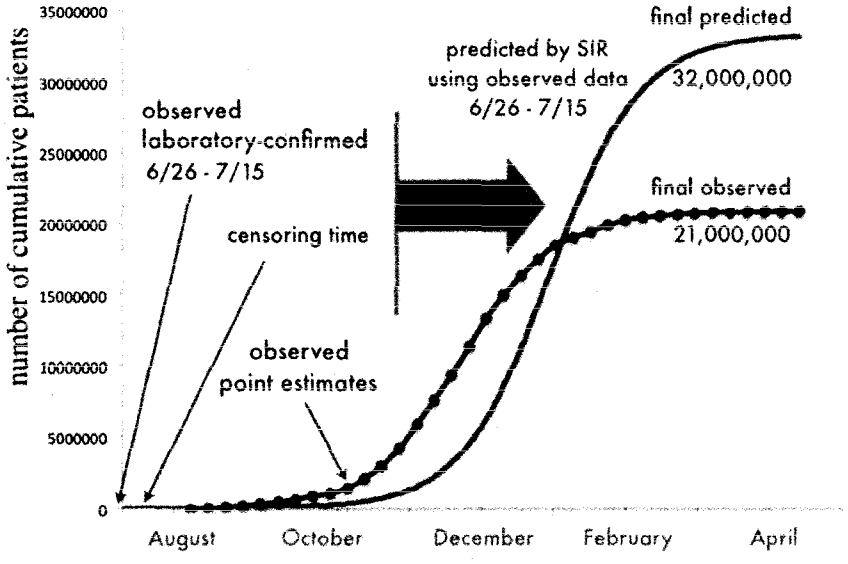

Fig. 11. Number of removed persons using the differential equations for Japan novel influenza case: final stage.

We can then proceed with the computation of the SEIR model. Computational results by the SEIR model are found to be very close to that of the MAS model. Therefore, we are convinced of a consistency between the models. Based on this consistency, we proposed to use a combined method of the MAS model and the SEIR model to estimate the final stage of the pandemic. This is the MADE simulation. By utilizing the MAS model in the early stage in a pandemic simulation to determine the appropriate parameters then applying

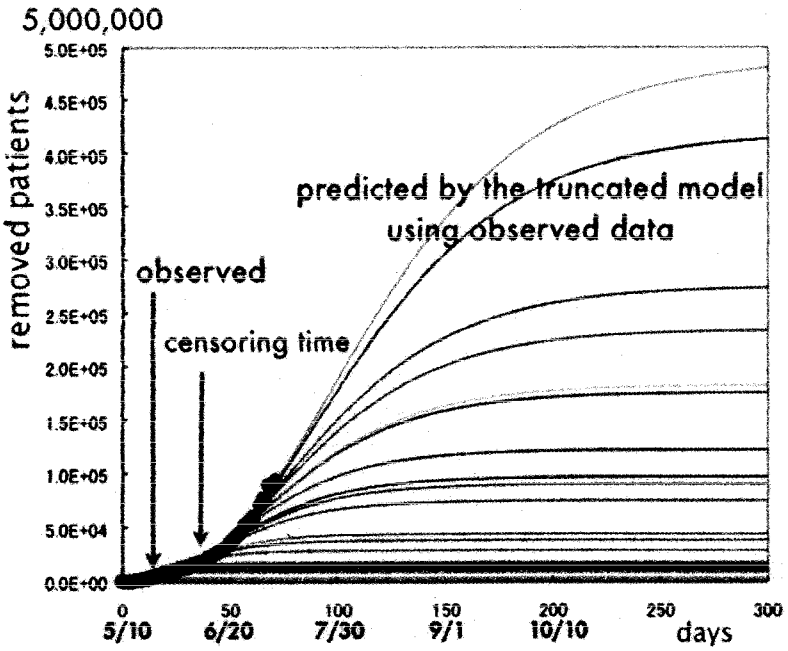

Fig. 12. Number of removed persons using the truncated model for Japan novel influenza case: early stage.

the SEIR model in the subsequent stage, with parameters obtained in the early stage using the difference equations, we can obtain the removed populations and other data with lesser computing cost and smaller computational errors.

Unlike the statistical parameter estimation results that provide incomplete data, the proposed approach is versatile enough for many situations. Although the parameters are obtained by the MAS even in early stages of pandemic, computational results for the final removed population size using the MADE are comparable to those calculated from the MAS model alone. With results by such a model, we can prepare for hazardous situations, such as an outbreak of new diseases or disease spread by acts of terrorism. The proposed method, the MADE, is superior to the statistical inference using the truncated model regarding the prediction accuracy.

Based on the similarity between the MAS and the SEIR results, we can trust the SIR results using the initial stage of the real pandemic. As an example case, we have discussed about prediction for the novel influenza spread. Although we cannot estimate the final stage accurately, a rough estimate can be obtained assuming that the infected persons will move to be removed persons in a few days. We have shown that the influenza have a possibility of large scale of spread in one year with very few observed data if no vaccination or human network shutdown is operated; after all, we have found that this prediction could be useful.

This paper focuses on the final stage prediction in pandemics. However, we should also pay attention continuously to other aspects; that is, an early detection such as [7] and [13] is another aspect. Some references, [17, 18, 20, 25], are directly related to that, and [19, 30] are are indirectly related to that. Disease propagation animation is seen in [40] for foot-and-mouth disease case in Japan and Korea.

7 Acknowledgement

The author appreciates Dr. Y. Toyosaka for his assistance. Since the paper was written on January 10, 2011, some references are added.

References

[1] R. Anderson and R. May, Infectious diseases of humans: Dynamics and control, Oxford University Press, 1991.
[2] C.L. Barrett, S.G. Eubank and J.P. Smith, If smallpox strikes Portland, Scientific American, 292 (2005) pp. 54-61.
[3] F. Brauer, P. van den Driessche and J. Wu (ed.), Mathematical Epidemiology, Lecture Notes in Mathematics, Springer, 2008.
[4] CDC, http://www.cdc.gov/H1N1FLU/
[5] CDC, http://www.cdc.gov/media/transcripts/2009/t090724.htm
[6] G. Chowell, A.L. Rivas, N.W. Hengartner, J.M. Hyman, and C. Castillo-Chavez, Critical response to post-outbreak vaccination against foot-and-mouth disease, Mathematical studies on human disease dynamics: emerging paradigms and challenges, AMS Contemporary Mathematics 47 (2006), 73-87.
[7] N.A. Christakis and J.H. Fowler, Social Network Sensors for Early Detection of Contagious Outbreaks, PLoS ONE, 5(9), (2010) 1-8.
[8] O. Diekmann and J.A.P. Heesterbeek, Mathematical epidemiology of infectious diseases: model building, analysis and interpretation, New York: Wiley, 2000.
[9] W.L. Deemer Jr., D.F. Votaw, Jr., Estimation of Parameters of Truncated or Censored Exponential Distributions, Ann. Math. Statist., 26 (1955), 498-504.
[10] L.R. Elveback, J.P. Fox, E. Ackerman, A. Langworthy, M. Boyd, L. Gatewood, An influenza simulation model for immunization studies, American Journal of Epidemiology, 103 (1976), 152-65.
[11] S. Eubank, Scalable, efficient epidemiological simulation, Proceedings of the 2002 ACM symposium on Applied computing, (2002) 139-145.
[12] N.M. Ferguson, C.A. Donnelly, R.M. Anderson, The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions, Science, 292 (11 May 2001) 1155-1160.
[13] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski & L. Brilliant, Detecting influenza epidemics using search engine query data, Nature 457 (2009) 1012-1014.

[14] H. Hirose, The mixed trunsored model with applications to SARS, Mathematics and Computers in Simulation, 74 (2007), 443-453.
[15] H. Hirose, Estimation for the size of fragile population in the trunsored and truncated models with application to the confidence interval for the case fatality ratio of SARS, Information, 12 (2009) 33-50.
[16] H, Hirose, K. Matsukuma, T. Sakumura, Infectious disease spread prediction models and consideration, Transactions of Information Processing Society of Japan - TOM, 4(3), (2011), 102-109.
[17] H, Hirose, Parameter estimation for the truncated Weibull model using the ordinary differential equation, Proceedings of Computers, Networks, Systems and Industrial Engineering, (2011) 396-399.
[18] H, Hirose, Estimation of the number of failures in the Weibull model using the ordinary differential equation, European Journal of Operational Research, 223(3), (2012) 722-731.
[19] H. Hirose, T. Nakazono, M. Tokunaga, T. Sakumura, S.M. Sumi, J. Sulaiman, Seasonal infectious disease spread prediction using matrix decomposition method, the 4th International Conference on Intelligent Systems, Modelling and Simulation, (2013)
[20] H, Hirose, L. Wang, Prediction of infectious disease spread using twitter: A case of influenza, Proceedings of the 5th International Symposium on Parallel Architectures, Algorithms and Programming, (2012) 17-20.
[21] C.Y. Huang, C.T. Sun, J.L. Hsieh, Y.M.A. Chen and H.L. Lin, A novel small-world model: Using social mirror identities for epidemic simulations, Simulation, 81 (2005) 671-699.
[22] 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, 141A (1933) 94-122.
[23] Y. Komori, H. Hirose, Parameter estimation based on grouped or continuous data fo truncated exponential distributions, Comm. Stat. - Theory and Method, 31 (2002), 889-900.
[24] I.M. Longini, Jr., M.E. Halloran, A. Nizam and Y. Yang, Containing pandemic influenza with antiviral agents, American Journal of Epidemiology,159 (2004) 623-633.
[25] Y. Maki, H. Hirose, Infectious disease spread analysis using stochastic differential equations for SIR model the 4th International Conference on Intelligent Systems, Modelling and Simulation, (2013)
[26] M.I. Metzer, N.J. Cox and K. Fukuda, The economic impact of pandemic influenza in the united states: priorities for intervention Emerging Infectious diseases, 5 (1999) 659-671.
[27] M. M. Mittal, R. C. Dahiya, Estimating the parameters of a truncated Weibull distribution Comm. Statist. - Theory Method, 18 (1989) 2027-2042.
[28] J.P. Narain, R. Kumer, R. Bhatia, Pandemic (H1N1) 2009: Epidemiological, clinical and prevention aspects, The National Medical Journal of India, 22 (2010) 1-6.
[29] F.J. Richards, A Flexible Growth Function for Empirical Use, Journal of Experimental Botany, 10 (1959) 290-301.
[30] S. Takimoto and H. Hirose, Recommendation systems and their preference prediction algorithms in a large-scale database, Information, 12(5) (2009) 1165-1182.
[31] Y. Toyosaka and H. Hirose, The consistency between the two kinds of pandemic simulations of the SEIR model and the MAS model, the 9th International Conference on Computers, Communications and Systems, (2008) 89-92.

[32] Y. Toyosaka and H. Hirose, Pandemic simulations by MADE: a combination of multi-agent and differential equations, The 2009 International Conference on Parallel and Distributed Processing Techniques and Applications (2009)331-335.
[33] Y. Toyosaka and H. Hirose, Pandemic Analysis in Swine Influenza A(H1N1), IEICE Technical Report, 109-232 (2009) 31-36.
[34] Y. Toyosaka and H. Hirose, The consistency of the pandemic simulations between the SEIR model and the MAS model, IEICE Transactions on Fundamentals, E92-A (2009) 1558-1562.
[35] Y. Toyosaka and H. Hirose, Pandemic Simulations by MADE: the Hybrid Method of Multi-Agent and Differential Equations, Asia Simulation Conference (2009) 1-5.
[36] WHO, http://www.who.int/csr/don/en/
[37] WHO, http://www.who.int/csr/sars/country/en/
[38] YouTube1, http://www.youtube.com/watch?v=NylT5leFHp4
[39] YouTube2, http://www.youtube.com/watch?v=6G-EaaIr8Ig
[40] YouTube3, http://www.significancemagazine.org/details/video/1337659/Foot-and-Mouth-Disease-Japan-v-Korea.html