Finding correspondence between agent-based and system dynamics models for infectous disease spread

Alexandre Vassalotti

McGill University, Montréal, Québec

Abstract

We present an approach for deriving a correspondence between the SIR epidemiological model to an equivalent agent-based model. We show a detailed correspondence analysis leading to an accurate fit between the two models.

Keywords: Epidemiology, SIR model, Agent-based modelling, Nonlinear dynamical systems

1. Introduction

Computational modelling of various properties of infectious disease outbreaks are an important tool in development of effective public health measures. For example, one can predict with accurate models the effectiveness of a public health measure in mitigating the spread of a disease. This is especially important since the high costs for implementating such measures often limits the amount of experimentation done in real world settings. Furthermore, ethical concerns limit the kinds of acceptable experimentation. Through simulation, we can make observations about the effectiveness of high risk and high cost measures, infect model populations with potent viruses and assist engineers and city planners to the design of metropolitan areas.

Despite their benefits, accurate infectous disease models are difficult to develop due to large number of parameters that needs to be accounted for. Distributions of age, sex and education, the topology of transportation networks and social structures, and the implementation of vacination campaigns, quarantines or other public health measures are only a few examples of parameters that has an influence on the spread of disease. To complicate things

Email address: alexandre.vassalotti2@mail.mcgill.ca (Alexandre Vassalotti)

further, many parameters have complex interrelationships which are difficult to quantify precisely.

To tackle this complexity, researchers reaches out in general to two formalisms: system dynamics and agent-based modelling. System dynamics describes the behaviour of a system over time using level and rate equations (accumulations and flows) structured into feedback loops (Forrester, 1994). The basis of this approach is the acknowledgment that the structure of system is as important as its individual components for explaining the behaviour of the system as a whole. Agent-based modeling is similar with that respect. However the two methodologies differ in the level at which they focus their attention and how they model the relationships among the components (Van Dyke Parunak et al., 1998). In agent-based models, we have a collection of automonous agents operating in bounded regions of a simulated world. Agents can display complex individual behaviour that changes over time, such as learning and adaption (Bonabeau, 2002). Furthermore, these agents can interact with one and another to create complex emergent behaviour. On the other hand, systems dynamics models typically assume homogeneous mixing and markovian behaviour for the individuals. Also, systems dynamics focus on aggregates rather individuals. Therefore, we can say a system modeled using an agent-based approaches captures a phenomena from the bottom-up where as system dynamics captures it from the top-down.

Finding a correspondence between two models is in general a diffcult task. Even for the simple case we present, we will have to settle for an approximation because of the Kermack-McKendrick model is not expressive enough to capture all the details of the agent-based. Despite this fundamental limitation, we will show that a good correspondence is possible. This is an interesting development because it shows that the complexity of an emergent phenomenon generated by an agent-based model can be captured by a simple system dynamics model. And this result could lead to more efficient simulation of such phenomenon by replacing computationally extensive agent-based models with computationally system dynamics models where applicable.

2. Kermack-McKendrick disease model

In epidemiology, an SIR model computes the theoretical number of infected people by a disease in a population over time t. It is characterized by the fact that it separates the population into three distinct subgroups (or compartments): the number of susceptibles people $S(t)$ that could be in-

fected with the disease, the number of infected people $I(t)$, and the number of people $R(t)$ that recovered from the disease. The first SIR models were proposed by Kermack and McKendrick (1927), but they gained prominence only decades later when they were brought back by Anderson et al. (1979). SIR models were shown to be good and simple models for understanding the dynamics of population under the influence of an infectious disease.

The Kermack-McKendrick model is perhaps the simplest example of SIR model. This model makes several simplifying assumptions about the population and the infectious disease:

- *Fixed population.* There is no births, deaths associated with the disease, or deaths of natural causes.
- *Homogeneous population.* Characteristics of the individuals, like age, sex, education and so forth, have no influence on the disease.
- *Instant incubation period.* The time period between an individual gets infected and becomes infectious is zero.
- *Disease imply infectivity.* The length of the infectious period is the same as the duration of the disease.
- *Permanent immunity.* Recovery from the disease confers permanent immunity to the individual.
- *Uniform diffusion.* The probability of an individual of being in contact with an infectious agent is the same as for every other individual.

While the assumptions of the Kermack-McKendrick model limit its applicability, it is a good choice for us because its simplicity.

In a fixed population $N = S(t) + I(t) + R(t)$, a possible system dynamics formulation of the Kermack-McKendrick model consists of three nonlinear coupled ordinary differential equations

$$\begin{cases} \dfrac{dS}{dt} = -\beta IS \\ \dfrac{dI}{dt} = \beta IS - \nu I \\ \dfrac{dR}{dt} = \nu I \end{cases}$$

This formulation implies the rate at which individuals acquire the infection is proportional to the number of contact between susceptible and ratio of infected individuals to the total population, where β is the transmission rate parameter. It also implies the rate at which individuals recover from the disease is proportional to number of infected, that is νI where ν is the recovery rate parameter.

Note, this model dictates the transition rate to the infected compartment, (as known as the force of infection) to be βI. This results in a force of infection that depends on the absolute number of agents in the population. So for a fixed β, the rate of infection will increase as we increase the population density N.

3. Agent-based disease model

To derive an agent-based model equivalent to Kermack-McKendrick model, we have to use all the same assumptions as previously and make some additional ones. One common characteristic of agent-based models is the notion of a domain, or world, in which the agents exists and interact in. To make the representation nice to work with, we choose the world W to be a continuous 2D space of unit area, or formally $\mathcal{W} \subseteq \mathbb{R}^2$ such that $\int_{x \in \mathcal{W}} dx = 1$. We also equip this world with an Euclidean distance metric $d(x, y)$ with the property that the area of influence $\mathcal{R} = \{y : d(x, y) < r,\ y \in \mathcal{W}\}$ defined around a given point $x \in \mathcal{W}$ and radius r satisfies $\int_{x \in \mathcal{R}} dx = \pi r^2$. This is our way of saying the world has no boundaries, or edges.

Then, we can define the behaviour of an individual agent as follow:

```
1  if State_t = Susceptible
2      for all agent in World
3          if agent.State_t = Infected and d(Position_t, agent.Position_t) < r
4              State_{t+1} ← Infected
5          else
6              State_{t+1} ← Susceptible
7  elseif State_t = Infected
8      if RANDOM(0, 1) < γ
9          State_{t+1} ← Recovered
10     else
11         State_{t+1} ← Infected
12 elseif State_t = Recovered
13     State_{t+1} ← Recovered
14 Position_{t+1} ← RANDOM-POINT(World)
```

Just to be clear, in the above rules, the procedure RANDOM-POINT($World$) returns a random point from a continuous uniform distribution over $\mathcal{W}$ and similarly RANDOM(0,1) returns a random value from a continuous uniform distribution over the closed interval $[0, 1]$. Also the radius of infection r and the probability of recovery γ are the parameters of our agents.

It is easy to check that the above rules respect all the assumptions we gave for the Kermack-McKendrick model. Uniform diffusion is the only non-obvious property. We achieve it using the random position update on line 14, which makes the agents "teleport" themselves around the world. Clearly this is not representative of how an individual would actually move in an environment, but we need this if we hope to find a correspondence with our systems dynamics model.

Another thing we may observe is we no longer have continuous time. The state and the position of an agent evolve in discrete time steps. Moreover, this model is not deterministic anymore. We now have a stochastic process. These are things we will need be aware during our correspondence analysis.

4. Correspondence analysis

Here, the goal of our correspondence analysis will be to find the system of equations between the parameters β and ν of the system dynamics model and the parameters r and γ of the agent-based model that will result in equivalent evolutions over time of $S(t)$, $I(t)$ and $R(t)$ under both models.

First, we find the trivial correspondence $\nu = \gamma$ for the recovery rates of the two models. Since the random values X returned by $\textsc{Random}(0, 1)$ have uniform distribution $X \sim \mathit{Uniform}(0, 1)$, it follows that the probability of recovery for an infectious subject is $\Pr(X < \gamma) = \gamma$.

Next, we want to derive the probability of an individual of being in contact with an infectious agent. So assuming an agent has a probability α of being in the infectious area $\alpha \leq 1$ of an infected agent in a world of unit area, then, the probability of of not being infected by k independent infectious agents is $(1 - \alpha)^k$. This leads to the probability of being infected by one of them, being $1 - (1 - \alpha)^k$.

This means the force of infection in the agent-based model is an exponential function in number of infected individuals I. However, in the Kermack-McKendrick model, we assumed it to be proportional to I. For this reason, we will need to get rid of the exponential and use an approximation to solve this model mismatch.

By equating forces of infection of the two models, we get

$$\beta I = 1 - (1 - \alpha)^I \tag{1}$$

Then, we can estimate β using a first order Taylor's expansion around $I = n$, being

$$\begin{aligned}
\beta &= \frac{1 - (1 - \alpha)^I}{I} && && (2)\\
&= \frac{1}{n} - \frac{1}{n} \exp\left(n \ln(1 - \alpha)\right) + \mathcal{O}(n - I) && \text{(around } I = n\text{)} && (3)\\
&= -\ln(1 - \alpha) + \mathcal{O}(I) && \text{(around } I = 0\text{)} && (4)
\end{aligned}$$

So which expansion should we choose? If the rate of recovery is zero (or sufficiently close), then we should take the expansion around $I = N/2$ because the effect of the infection rate β will be the greatest around that point. In other words, when $\nu = 0$ in the systems dynamics model, the equation

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

is maximized when $I = N/2$. Hence, it would make sense to have a more accurate approximation around that region to mitigate the effects of the error term. We can get the correspondence for this case by substituting α for the

area of infection πr^2

$$\beta \approx \frac{2}{N} - \frac{2}{N} \exp\left(\frac{N}{2} \ln(1 - \pi r^2)\right) \tag{6}$$

Unfortunately, finding the optimal point to expand around is a difficult problem in general due to the non-linearity of the model. A simple heuristic is to look at the basic reproduction ratio $R_0 = N\beta/\nu$. When R_0 is small, let say $R_0 < 3$, the infection will die out quickly and we can approximate β with the expansion around $I = 0$ to get reasonable results.

A better heuristic however is to expand at the maximum of I. So if we solve $dI/dt = 0$, then we find the maximum occurs when $S = \nu/\beta$. And by integrating dI/dS, we get

$$I = S_0 + I_0 - \frac{\nu}{\beta} \ln S_0 + \frac{\nu}{\beta} \ln S - S \tag{7}$$

where S_0 and I_0 are the initial numbers of susceptible and infected individuals. Plugging in $S = \nu/\beta$ gives

$$I = S_0 + I_0 + \frac{\nu}{\beta}\left(\ln S_0 - \ln \frac{\nu}{\beta} + 1\right) = \eta(\beta) \tag{8}$$

Since the maximum depends on β, we need to use an iterative scheme to get our approximation. Assuming $r > 0$ and $\nu > 0$, then we can combine (3), (4) and (8) to the estimate for β

$$\beta_i = \frac{1}{\eta(\beta_{i-1})}\left(1 - \exp\left(\eta(\beta_{i-1}) \ln(1 - \pi r^2)\right)\right) \tag{9}$$

$$\beta_0 = -\ln(1 - \pi r^2) \tag{10}$$

When $\nu = 0$, we will use our previous approximation (6). And if $r = 0$, then $\beta = 0$. This scheme will converge linearly toward a desired estimate of β as $i \to \infty$. In practice, it only takes a few iterations to achieve convergence within a machine epsilon.

5. Results

To assess the quality of our correspondence, we implemented simulators in Python for the two models presented and ran simulation for multiple

combinations of parameters. We found the correspondence to be especially good when the infection is strong. We also found the iterative scheme to give consistently good results. We present sample results in Figure 5 and 3. But unfortunately, due to the runtime of the simulations, we are unable to provide statistical analysis of our results.

6. Conclusion

Even for a simple model, like the Kermack-McKendrick model, we shown that achieving good correspondences between models constructed with different formalisms is difficult but possible. The approach we presented is flexible enough to be applied to other kinds of infectious disease models.

Our analysis could be improved however by looking more in depth at the differences of specification between agent-based and system dynamics models, such as the effects of discrete versus continuous time and stochastic versus deterministic representations. For the model we presented, our analysis was sufficient but other models these difference could have a great impact on how we derive correspondences.

Finally, it would be instructive to attempt to derive correspondences to agent-based models with other characteristics than the one we presented to understand better the limitations of the Kermack-McKendrick model, and other SIR models. For example, we could use discrete space representations, such as grids and graphs, instead of a continuous area. Or, we could attempt to deliberately break some of the assumptions of the Kermack-McKendrick to analyze its robustness. This further research could result in better understanding about the applicability of such models.

References

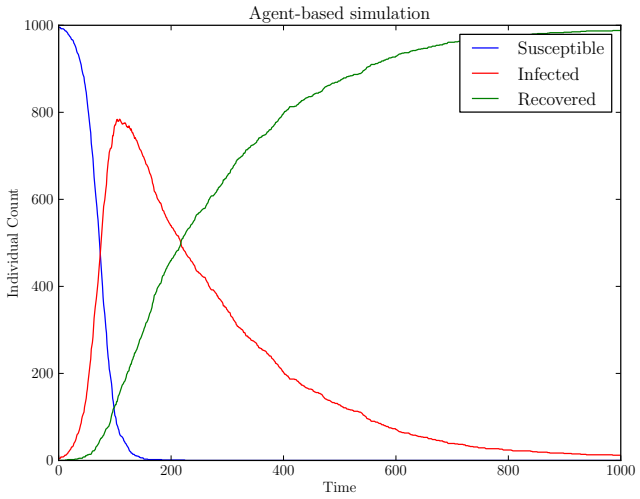
Anderson, R., May, R., et al., 1979. Population biology of infectious diseases: Part i. Nature 280 (5721), 361.

Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 99 (Suppl 3), 7280.

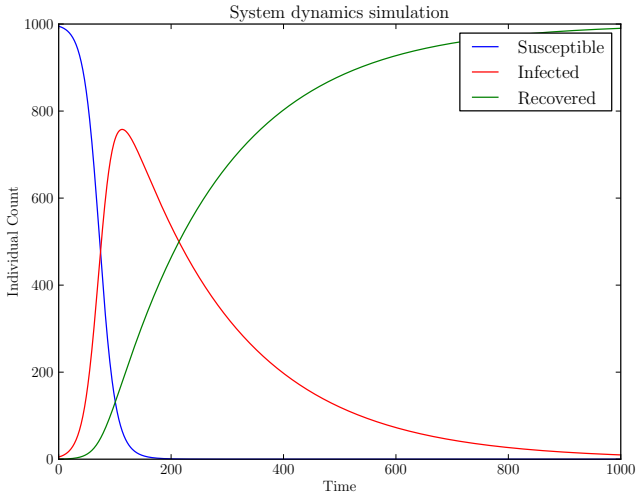
Forrester, J., 1994. System dynamics, systems thinking, and soft OR. System Dynamics Review 10 (2-3), 245–256.

Kermack, W. O., McKendrick, A. G., 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A 115 (772), 700–721.

Van Dyke Parunak, H., Savit, R., Riolo, R., 1998. Agent-based modeling vs. equation-based modeling: A case study and users guide. In: Multi-Agent Systems and Agent-Based Simulation. Springer, pp. 277–283.



(a) Result of a simulation of the agent-based model with radius of infection $r = 0.005$, and rate of recovery $\gamma = 0.005$



(b) Result of a corresponding simulation of the system dynamic model, rate of infection $\beta = 7.62562 \times 10^{-5}$ and rate of recovery $\nu = 0.005$

Figure 1: Simulation results for the evolution of a population of size $N = 1\,000$ over $1\,000$ days under the influence of an infectious disease. The parameter β was estimated using (10).

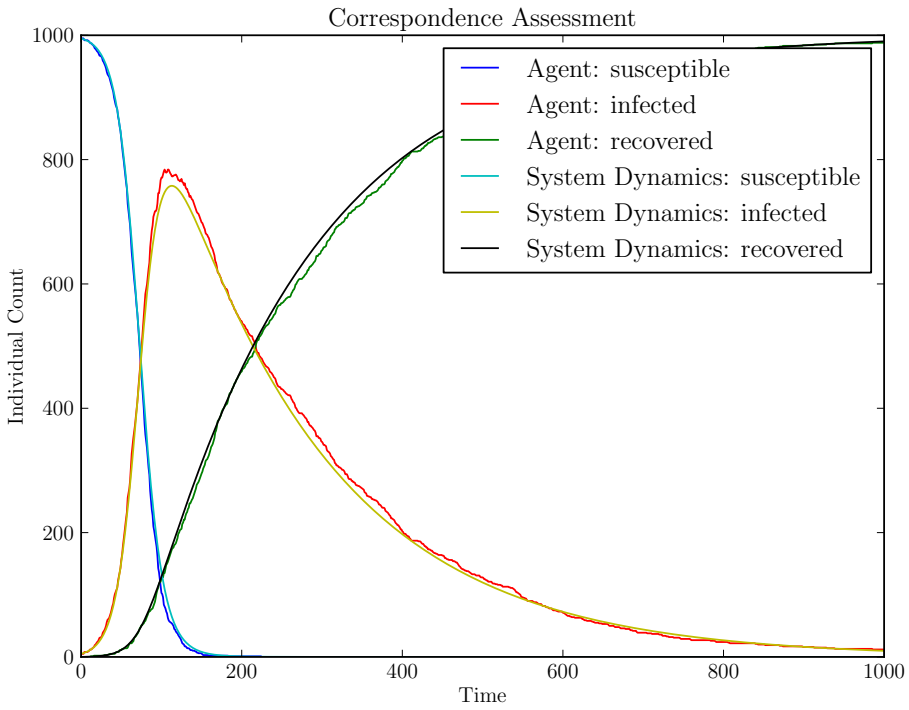
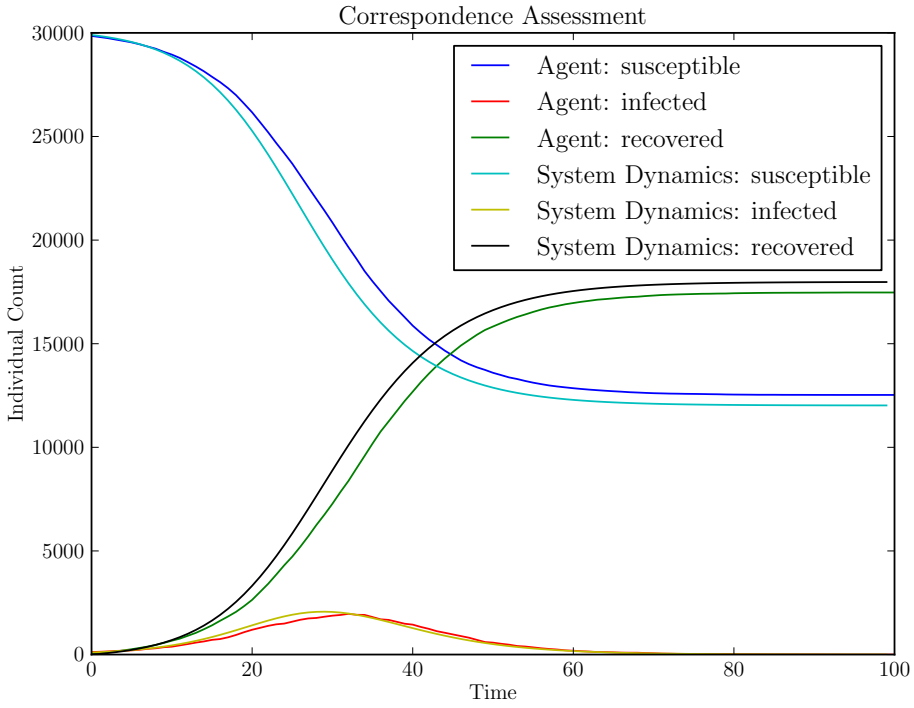
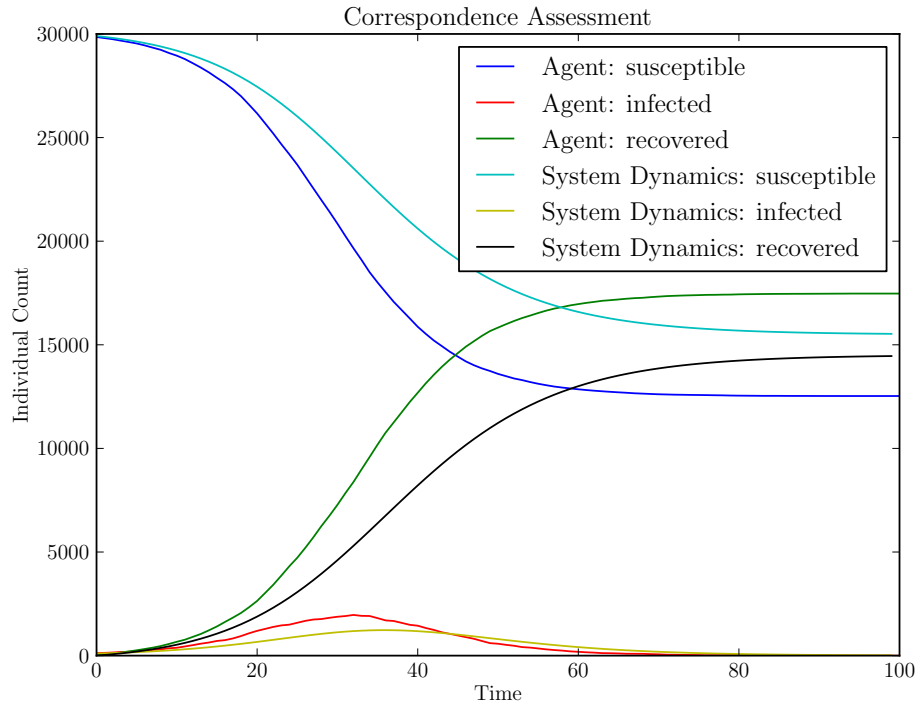


Figure 2: Comparison of the simulation results from Figure 1. RMS of the susceptible, infected and recovered differences between the two simulations are 4.71, 10.62 and 8.63 respectively



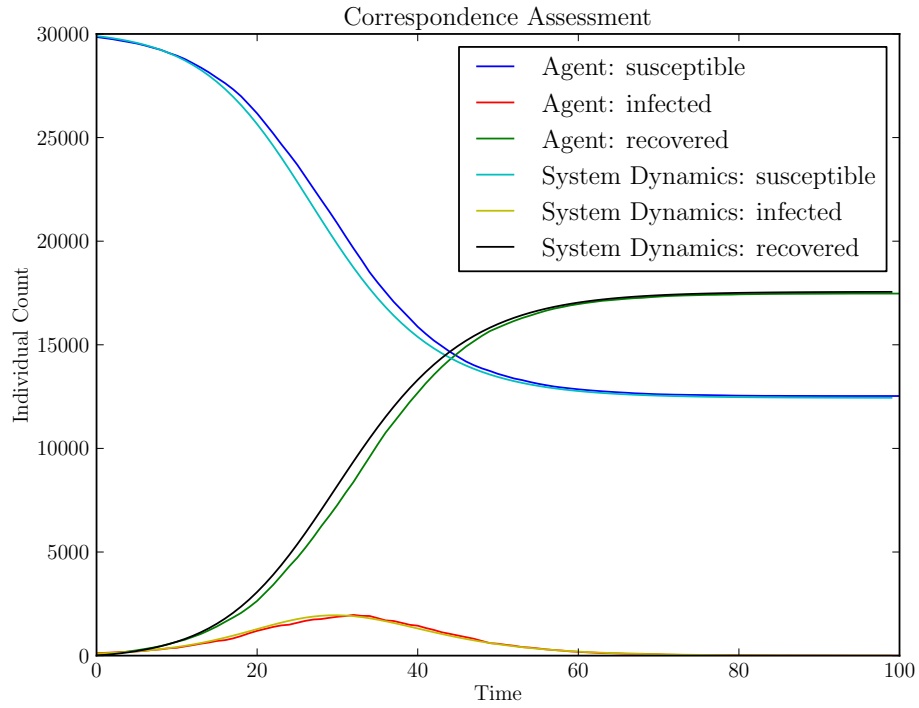
(a) Result with the parameter $\beta = 1.5205 \times 10^{-5}$ for the system dynamics model obtained using the approximation around $I = 0$. RMS of the susceptible, infected and recovered differences between the two simulations are 838.7, 116.3 and 813.2 respectively.

Figure 3: Comparison of the approximations around $I = 0$, $I = N/2$ and $I = \eta$ of β in a simulation of a population of size to $N = 30\,000$ and $I_0 = 100$ over $1\,00$ days under the influence of a weak infection. The parameter of the agent-based model are $r = 0.0022$ and $\gamma = 0.3$.



(b) Result with the parameter $\beta = 1.3596 \times 10^{-5}$ for the system dynamics model obtained using the approximation (6) around $I = N/2$. RMS of the susceptible, infected and recovered differences between the two simulations are 3219.2, 317.7 and 3174.0 respectively.

Figure 3: (continued) See page 12 for the description.



(c) Result with the parameter $\beta = 1.4982 \times 10^{-5}$ for the system dynamics model obtained using the approximation (10) around $I = \eta$. RMS of the susceptible, infected and recovered differences between the two simulations are 381.9, 63.9 and 370.0 respectively.

Figure 3: (continued) See page 12 for the description.