A Simple Efficient Technique to Adjust Time Step Size in a Stochastic Discrete Time Agent-based Simulation

Chia-Tung Kuo, Da-Wei Wang and Tsan-sheng Hsu

Institute of Information Science, Academia Sinica, Taipei, Taiwan

Keywords: Simulation, Step Size, Efficiency, Granularity.

Abstract: This paper presents a conceptually simple approach on adjusting the time step size in a stochastic discrete time agent-based simulation and demonstrates how this could be done in practical implementation. The choice of time step size in such a system is often based on the nature of the phenomenon to be modelled and the tolerated simulation time. A finer time scale may be desired upon the introduction of new events which could possibly change the system state in smaller time intervals. Our approach divides each original time step into any integral number of equally spaced sub-steps based on simple assumptions, and thus allows a simulation system to incorporate such events and produce results with finer time scale. Regarding the trade-off between finer scale and higher use of resource, our approach also highlights the implementation techniques that increase the resource usage and simulation time only marginally. We analyze the results of this refinement on a stochastic simulation model for epidemic spread and compare the results with the original system without refinement.

1 INTRODUCTION

The stochastic discrete time simulation model is a useful and efficient way in simulating agent-based activities and has gained significant popularity in modelling many dynamical biological or physical systems in recent years, such as the dynamics of epidemic spread (Tsai et al., 2010a; Riley, 2007). Often time, such simulations can give insights into problems where traditional models are too complicated and analytic results are very difficult or currently impossible to obtain (Tsai et al., 2010a; Germann et al., 2006). This high level abstraction, however, introduces artifacts which do not pertain to real world behaviour, namely, the discretization of time. In particular, event simultaneity whereby multiple distinct events occur at exactly the same time may be due to an insufficiently detailed discrete-time model (Vangheluwe, 2001). The additional parameter, the size of the time step, can potentially have significant impact on the results of simulation without the awareness of the modeller (Buss and Rowaei, 2010). Buss and Rowaei investigated time advancement mechanism and the role of time step size in a somewhat different context (Buss and Rowaei, 2010; Buss and Rowaei, 2011), and found no systematic studies had been done with its effects. Moreover, the choice of time step si-

ze plays a role in the efficiency of the simulation as the occurrence of events need to be checked more frequently for smaller time steps. Accordingly, the size of the time step should be carefully selected to match the real world phenomenon to be modelled as realistically as possible, and without too much sacrifice of simulation time. Other than for its realistic nature, this choice of time step size is often influenced by the empirical data we have. For example, in modelling the epidemic spread, the unit of measurements for latent and infectious periods also affect the choice of time step size (Kelker, 1973). Specifically, the unit of measurements for latent and infectious periods should be smaller than or equal to the length of a time step to make sure that no event advances two steps in one simulated time step.

We have developed a simple, yet efficient, technique based on reasonable assumptions to split each time step into any integral number of equally spaced sub-steps with small increase in simulation time in a stochastic discrete time agent based simulation system for epidemic spread we developed earlier (Tsai et al., 2010a; Tsai et al., 2010b). More precisely, we first modify the probabilities of events according to the basic probability theory such that they correspond to events occurring in a smaller time frame. This allows the introduction of certain types of events

in a smaller time scale. Then we introduce a structurally different implementation model that significantly outperforms the straightforward step-by-step linear model. The experiments show that our improved implementation produces stochastically identical results as the straightforward implementation with significantly less time. Furthermore, our results also show that, given the same set of possible events, simulation with a finer time scale causes events to occur slightly earlier. This is well justified since we assign certain probability of occurrence to each event in a smaller time interval and update the system state more frequently.

This paper is organized as follows. Section 2 gives a simplified view of the basic model of a stochastic discrete time agent-based simulation. Section 3 describes our adjustment and its implementation. Section 4 demonstrates our experimental results and provides a discussion of these results. Section 5 describes the general applicability of our approach and discusses its limitations. Section 6 concludes our paper and points out directions for future work.

2 SIMULATION MODEL

In this section, we describe the basic model of a stochastic discrete time agent-based simulation system, with specific reference to the simulation system developed by (Tsai et al., 2010a) following approach in (Germann et al., 2006). Algorithm 1 provides a high level description of the agent-based simulation model of an epidemic spread.

Algorithm 1: Stochastic agent-based simulation model.

1: **for each** time step T from *beginning* to *end_of_simulation* **do**
2:     **for each** infectious agent I **do**
3:         **for each** susceptible agent S in contact with I during T **do**
4:             **if** I infects S successfully **then**
5:                 update status of S
6:             **end if**
7:         **end for**
8:     **end for**
9: **end for**

The time step refers to the indexing variable in the outermost loop. This step size characterizes the unit of time in which the system progresses. In other words, the system state must remain constant between successive steps; thus, all events must occur with periods of integral multiples of this time step size. In line 2, we identify the agents that may change the system state in the current time step. In the inner loop, for each of these agents, we find all interacting agents in the same step, and decide whether an event between agents actually occurs. There are often times scheduled events other than among interacting agents that may take place; these could be dealt with similarly and we focus our attention on events described in Algorithm 1.

In our simulation of an epidemic spread, each time step is one half-day. This is largely due to the facts that agents (people) are in contact with different other agents during daytime and nighttime, and also state transitions take place in multiple of half-days. An agent is classified as one of the susceptible, infectious, and recovered (or removed); in each step, an infectious agent may infect susceptible agents in contact according to a transmission probability, which is dependent upon the interacting agents' ages and contact locations. Once an infectious agent successfully infects a susceptible agent, the susceptible will be assigned a latent period and an infectious period, and become infectious at the end of the half-day step. These latent and infectious periods are drawn from two pre-specified discrete random variables, derived from observed data on the epidemic to be modelled.

Besides the obvious possibility where finer time scale result is desired, there are still at least two potential issues with this current model regarding the size of the time step. First, if we wish to model a disease for which the latent and infectious periods are better modelled with finer time unit, the current system could not easily accommodate this change without major revision efforts. Second, if we wish to (indeed we do) record not only who is infected, but also the infector, then the artificial simultaneity (See Section 1) may be introduced: two or more infectious nodes may infect the same susceptible agent in one step, and thus some mechanism is required to decide the infector precisely.

In the next section, we introduce a modified model, which is the same as the original model in principle but divides each step into integral number of smaller sub-steps. In section 3.2, we describe efficient techniques in implementation that could achieve our desired results with minimal increase in simulation time.

3 ADJUSTMENT OF TIME STEP SIZE

3.1 Finer Step Model

In the original model, each pair of infectious (I) and susceptible (S) in contact are associated with a transmission probability P^{IS}. In each half-day step, we iterate through all pairs of infectious and susceptible in contact, and for each pair, decide whether S is infected by I with probability P^{IS}. Now, for each P^{IS}, we derive a k-hour transmission probability, p_k^{IS}, that satisfies

$$(1 - p_k^{IS})^{\frac{12}{k}} = 1 - P^{IS} \qquad (1)$$

where k is a factor of 12 (we use the term, *granularity* of the system, to denote the smallest unit of time interval in which an event to be modelled could take place). The probability p_k^{IS} is derived such that the overall probability of S getting infected by I does not change if S is decided for infection with probability p_k^{IS} every k hour(s), provided that no change in state occurs in each half-day step. Notice that this derivation also makes the assumptions that the probability of transmitting a disease is uniform in the half-day step and independent among each smaller sub-steps. That is, we use the same p_k^{IS} for all sub-steps, instead of a number of different (conditional) probabilities. These may be debatable assumptions, depending upon what events are being modelled.

The probability p_k^{IS} is derived between each *IS*-pair; we would like the probabilities of transmitting a disease (between any pair) in each step to be the same as the original probability, P^{IS} in cases of multiple pairs of infectious and susceptible agents in contact. Now we show that this is indeed the case. More formally, the probability of each susceptible agent S getting infected remains the same as long as the duration of contact between each *IS*-pair is unchanged and a multiple of 12-hour. Notice that it suffices to demonstrate the case where there are more than one infectious agents in contact with only one susceptible agent, as susceptible agents do not influence each other. This is an immediate result from the assumed independence of infection events by different infectious agents and the commutative property of multiplication. Suppose there are n infectious agents, $I_1, ..., I_n$ and one susceptible agent S in contact in some arbitrary half-day step. Let S_t and S^t denote the events S gets infected at t and S gets infected by t, respectively. Also, we use $\neg$ to denote logical negation. Then,

$$\begin{aligned} \Pr\{S \text{ infected}\} &= 1 - \Pr\{S \text{ not infected}\} \\ &= 1 - \Pi_{t_j = jk,\, j \in \mathbb{N}}^{12} \Pr\{\neg S_{t_j} | \neg S^{t_{j-1}}\} \\ &= 1 - \Pi_{t_j = jk\, j \in \mathbb{N}}^{12} \Pi_{i=1}^{n} \Pr\{\neg S_{t_j} \text{ by } I_i | \neg S^{t_{j-1}}\} \\ &= 1 - \Pi_{i=1}^{n} \Pi_{t_j = jk,\, j \in \mathbb{N}}^{12} (1 - p_k^{I_i S}) \\ &= 1 - \Pi_{i=1}^{n} (1 - P^{I_i S}) \end{aligned}$$

Notice that in the derivation above, we split one half-day into $12/k$ sub-steps of k hour(s) each. The same approach could be used for splitting a time step of any size into any integral number of equally spaced sub-steps.

This refinement does not introduce any conceptually new artefact into the model. All it does is to perform the simulation with shorter time step size, and in each time step, the probabilities for events to occur are altered. Specifically, in Algorithm 1, we substitute a step with a smaller sub-step in the outermost loop, and use p_k^{IS} instead of P^{IS} when deciding infection (line 4). Regarding the two concerns we have at the end section 2, the first is solved as we can now model any events that take place with periods greater than or equal to the *granularity* of the system (k hours in this example). This approach does not deal with the second concern directly. However, by reducing the time step size (and thus the transmission probability in a step), the chances of simultaneous events could be reduced significantly.

Now a new issue concerning efficiency is introduced. Typically, in a large scale agent-based simulation system, the number of possible interactions among agents or other events (line 2 in Algorithm 1) is very large in each step. Therefore, after applying this technique to reduce the step size, we will examine, in each step, a long list of possible events, of which most will not take place due to the reduced probabilities. In response to this efficiency issue, in section 3.2 we introduce techniques in implementation, which allow the system to run almost as fast as with the coarser time step, but achieve the benefits produced by the finer time step.

3.2 Efficient Implementation

Our goal in this section is to implement the refined model more efficiently. For the ease of description, we refer to the original time step (e.g. half-day in the model above) as *step*, and the finer time step (e.g. k hour(s)) as *sub-step*, and also maintain the use of P and p_k to denote the transmission probabilities in each step and sub-step, respectively. The challenge is that we wish to achieve the effect of advancing the system every sub-step unit of time, but we do not want to examine all possible transmission events such fre-
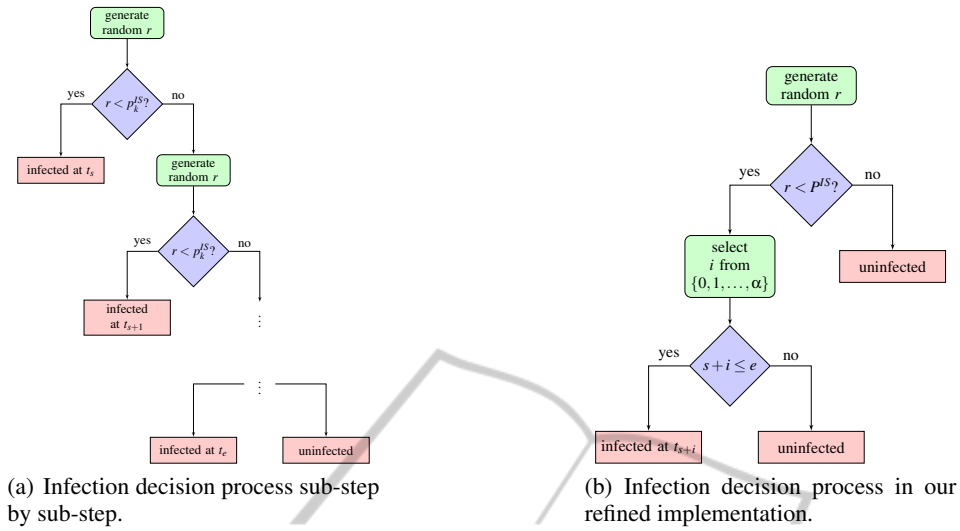
(a) Infection decision process sub-step by sub-step.

(b) Infection decision process in our refined implementation.

Figure 1: Schematic pictures of infection decision processes for an IS-pair in one full step T; step T consists of sub-steps t_1, t_2, ..., t_α, and I starts infecting at $t_s \in T$ and recovers at $t_e \in T$ where t_e is later than t_s. Notice that in both (a) and (b), r is drawn uniformly random from $(0, 1)$, while in (b) i is drawn from the computed distribution accordingly.

quently. The main idea here is to make the "big decision" first, and make subsequent "small decisions" only if the first result turns out "favorable". First, we "select" a possible transmission event with the aggregate probability, P^{IS}. This determines whether this agent gets infected in a step. Following that, we decide which sub-step this event actually occurs according to probabilities, $\frac{p_k^{IS}}{P^{IS}}$, $\frac{p_k^{IS}(1-p_k^{IS})}{P^{IS}}$, $\frac{p_k^{IS}(1-p_k^{IS})^2}{P^{IS}}$, ... Flow charts of infection decisions between an IS-pair are shown in Figure 1.

From Figure 1, it is straightforward to verify that in our modified implementation, the probability of an infectious agent transmitting the disease to a susceptible agent in contact in a step is the same as in the original model. In cases of multiple IS-pairs, an argument similar to the one in Section 3.1 could show that the probabilities of transmission do not change, provided the length of duration for any IS-pair in contact is fixed. There is, however, one more premise for this implementation to work: the contacts among agents in a step must be known before we start the step. This is necessary since we iterate through infectious agents only every step, instead of every sub-step as in Section 3.1. This assumption is often easily satisfied as in many agent-based simulations, the possible interactions among agents are determined by the pre-initialized properties of the agents (e.g. the contact locations in which agents reside).

The argument above shows that this improved implementation should, in principle, achieve the same result as the straightforward sub-step by sub-step implementation (Algorithm 1). However, we point out

three issues (may be more for more complicated models) for which extra care should be taken in practical implementation, in order to get this expected result:

1. Notice in the actual model with period being a sub-step, an event may become ready or may be removed between two successive sub-steps within a step. For example, in the model of epidemic spread, an infectious agent I may recover, or a susceptible agent S may turn infectious in any sub-step, so that contact with I no longer results in new infections, or new transmission routes become possible within a step. In these cases, assuming the knowledge of when events are ready or removed is known before each step, we can perform a simple lookup and filter out those events that we have determined its occurrence in a sub-step before it is ready or after it has been removed.

2. It is possible that two events could each occur with some probabilities, but they could not occur both in the same sub-step. That is, the occurrence of one prevents the occurrence of the other one. This issue pertains to the problem of "artificial simultaneity" in section 1 where some arbitrary ordering is needed. For example, a susceptible agent S cannot get infected from two infectious agents I_1 and I_2 in two different sub-steps of the same step. To conform to the original model (Algorithm 1) with time period being a sub-step, we must ignore the occurrence of all conflicting events in a step except the one which occurs in the earliest sub-step (There could still be more than one occurring in the earliest sub-step; in this case, an ar-

bitrary selection is made. But the chances of such cases are significantly reduced as indicated in section 3.1).

3. The occurrence of an event in a sub-step may introduce new events that are possible to occur immediately starting from the next sub-step. To deal with these cases, a list of these possible new events must be maintained in a step, and each event in this list is to be decided for its occurrence and cleared in the current step. This iterative examination will eventually terminate as fewer events will be added for the remaining sub-steps and probabilities of occurrence are smaller as well. If a later examined event $\mathcal{A}$ takes place and prevents the occurrence of an earlier decided event $\mathcal{B}$, which occurs later than $\mathcal{A}$ in time (as illustrated in issue 2), we must update the system accordingly to take event $\mathcal{A}$ and ignore event $\mathcal{B}$. For example, consider the case where in some step T, an infectious agent I_1 successfully infects a susceptible agent S_1 at sub-step t_1 and another susceptible agent S_2 at a later sub-step t_2 ($t_1 < t_2$, but they belong to the same step T). Assume S_1 turns infectious immediately and successfully infects S_2 in sub-step t_3 where $t_1 < t_3 < t_2$. Then we must update the infection time of S_2 to be t_3 instead of t_2. In practical implementation, the list of these newly triggered events within the same step should be sorted in order of the sub-steps of occurrence to avoid a long sequence of updates in the cases of the example above.

Algorithm 2 gives a high-level description of the practical implementation of the system that could achieve the same results as in Algorithm 1 with finer step size.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we wish to demonstrate that our refined model indeed achieves stochastically identical results as the original model with a shorter time period, and significantly reduces the simulation time. Also, we give a reasonable account for the observed differences in results from simulations of different time step sizes.

We build the two simulation systems (denoted by ALG_1 and ALG_2 for Algorithm 1 and Algorithm 2, respectively) by modifying a simulation system for epidemic spread developed by (Tsai et al., 2010a), keeping all parameters as original except the ones we explicitly wish to manipulate. Below, we perform 100

Algorithm 2: Refined stochastic agent-based simulation model.

```
 1: for each time step T from beginning to
       end_of_simulation do
 2:    initialize an empty sorted list L = ∅
 3:    for each infectious agent I do
 4:       TryToInfect(I,T,L)
 5:    end for
 6:    while L is not empty do
 7:       I_new ← remove the head of L
 8:       TryToInfect(I_new,T,L)
 9:    end while
10: end for

11: procedure TRYTOINFECT(I,T,L)
12:    for each susceptible agent S in contact with I
          during T do
13:       if I is still infectious and infects S success-
             fully in sub-step t, and S has not been infected
             before t then
14:          update status of S
15:          if S turns infectious within the current
                step T then
16:             Add S to L
17:          end if
18:       end if
19:    end for
20: end procedure
```

---

baseline simulations for each system with a particular granularity, and report the average results for both the simulation outputs and the simulation time consumed.

Figure 2(a) shows the epidemic curves (daily new cases) produced by the two systems, ALG_1 (blue) and ALG_2 (red), when granularity (GR) is set to 1 (+), 6 (o), and 12 (x) hour(s), respectively. Each pair of curves corresponding to the same granularity overlap well, showing that this modified implementation indeed reproduces the result of the original model. Moreover, the leftward shift of epidemic curves for finer granularities is also expected. This is a result of our assumption that the probability of transmission in each smaller step is uniformly distributed, and this causes the expectation of infection time to become earlier. Similar phenomenon is also observed if we shorten the latent period since it does not affect infectivity while causing a earlier spread of disease. Figure 2(b) shows the epidemic curves produced by the original system with varied latent periods.

Table 1 shows the average simulation times (Sim-Time in seconds) and attack rates (AR) for each of the two systems run on a workstation with 8 Intel Xeon X5365 CPUs and 32GB RAM. The consistency in at-
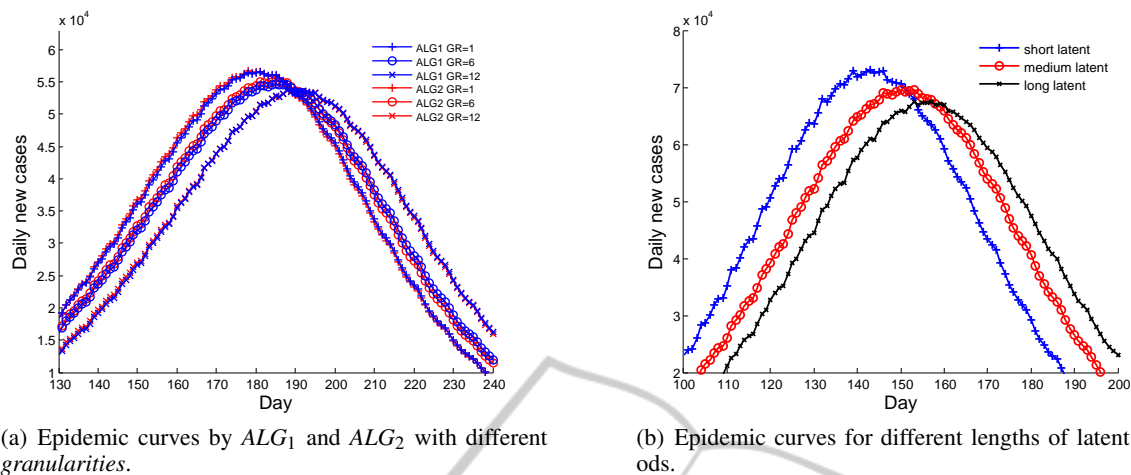
(a) Epidemic curves by ALG_1 and ALG_2 with different *granularities*.

(b) Epidemic curves for different lengths of latent periods.

Figure 2: Average epidemic curves from 100 simulations.

Table 1: Average attack rates and simulation times for the two systems ALG_1 and ALG_2 with different granularities.

GR (hr)	AR (ALG_1)	AR (ALG_2)	SimTime (ALG_1)	SimTime (ALG_2)
1	0.312	0.312	971	134
2	0.312	0.312	504	133
3	0.312	0.312	351	133
4	0.312	0.312	276	133
6	0.312	0.312	203	133
12	0.311	0.311	127	130

tack rates confirms our argument that the overall probabilities of infection are not altered; the huge reduction in simulation time demonstrates the effectiveness of our proposed model for practical implementation.

5 APPLICABILITY AND LIMITATIONS

We briefly describe the context in which this approach is applicable, as well as its current limitations. This approach was initially designed to handle situations in epidemic spread where we want a disease's natural history that is finer than the time step size in the original model. This was achieved by certain assumptions and basic probability theory, as shown in section 2. Algorithm 2 was then developed to improve the efficiency of the straightforward model (Algorithm 1) with a small step size and small interacting probabilities among agents. As a result, it should be easily applicable to other agent-based simulations when a similar factor affecting agents' interactions needs to be measured in finer time scale; that is agents could become active (e.g. infectious) or inactive (e.g. recovered or isolated) in such time scale.

It is crucial to notice that the sub-step in which

agents become inactive must be known before the start of each step (the case of turning active in the middle of a step could be handled as described in section 3.2 issue 3). It is not a problem in our example of epidemic spread simulation as the length of latent and infectious periods of an agent (thus the information of when it recovers) is determined when the agent gets infected. This, however, may pose a problem for other kinds of simulations. It may require tracing back of when an agent becomes inactive, and undoing all its interactions thereafter.

6 CONCLUDING REMARKS AND FUTURE WORK

We proposed a simple approach in adjusting the time step size in a stochastic discrete time agent-based simulation model based on reasonable assumptions. This provides flexibility in modelling events with finer time scales; such flexibility is often desired when finer empirical observations were to be incorporated into simulations. Furthermore, we described a structurally different model for practical implementation, which achieves identical result significantly faster. To demonstrate this, we modified a simulation system for

epidemic spread, developed by (Tsai et al., 2010a), constructed the proposed simulation model in both ways (Algorithm 1 and Algorithm 2), and through experiments, showed that they indeed computed the same results with the later one having a huge reduction in simulation time.

The implementation technique introduced in this paper may also be applied to purely increase the efficiency without any effect on results; we could view the original time step as the *sub-step* and run the system with an enlarged *step* following Algorithm 2 to produce the same result more efficiently. This may, however, introduce difficulties in determining the possible events to occur in an enlarged step. In the example of the epidemic spread simulation, two agents may be in contact only in daytime and not nighttime (see Algorithm 1 line 3); in an enlarged step, attention must be paid to such circumstances. It is also likely that other calibrated parameters need to be rescaled appropriately when such refinement is employed. We will work on overcoming the limitations mentioned above and in section 5, and try to apply such techniques to a larger variety of general simulations.

Another interesting direction is to compare the effect of time step size in discrete time agent-based simulations with the more traditional approaches to simulation modelling, such as event scheduling, and also the modelling with differential equations, commonly seen in mathematical epidemiology (Diekmann and Heesterbeek, 2000).

ACKNOWLEDGEMENTS

REFERENCES

Buss, A. and Rowaei, A. A. (2010). A comparison of the accuracy of discrete event and discrete time. In *Proceedings of the 2010 Winter Simulation Conference*.

Buss, A. and Rowaei, A. A. (2011). The effects of time advance mechanism on simple agent behaviors in combat simulations. In *Proceedings of the 2011 Winter Simulation Conference*.

Diekmann, O. and Heesterbeek, J. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis, and Interpretation*. John Wiley and Sons Ltd.

Germann, T. C., Kadau, K., Ira M. Longini, J., and Macken, C. A. (2006). Mitigation strategies for pandemic in-

fluenza in the united states. In *Proceedings of the National Academy of Sciences*, volume 103, pages 5935–5940.

Kelker, D. (1973). A random walk epidemic simulation. *Journal of the American Statistical Association*, 68(344):821–823.

Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316(5627):1298–1301.

Tsai, M.-T., Chern, T.-C., Chuang, J.-H., Hsueh, C.-W., Kuo, H.-S., Liau, C.-J., Riley, S., Shen, B.-J., Shen, C.-H., Wang, D.-W., and Hsu, T.-S. (2010a). Efficient simulation of the spatial transmission dynamics of influenza. *PloS ONE*.

Tsai, M.-T., Wang, D.-W., Liau, C.-J., and sheng Hsu, T. (2010b). Heterogeneous subset sampling. In *Lecture Notes in Computer Science*, volume 6196, pages 500–509.

Vangheluwe, H. (2001). Discrete event modelling and simulation. *Lecture Notes, CS522 McGill University*.