Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

A HYBRID EPIDEMIC MODEL: COMBINING THE ADVANTAGES OF AGENT-BASED AND EQUATION-BASED APPROACHES

Georgiy V. Bobashev
D. Michael Goedecke
Feng Yu

Statistics and Epidemiology Division
RTI International, RTP NC, 27709, U.S.A.

Joshua M. Epstein

Center on Social and Economic Dynamics
Brookings Institution,
Washington DC 20036, U.S.A.

ABSTRACT

Agent-based models (ABMs) are powerful in describing structured epidemiological processes involving human behavior and local interaction. The joint behavior of the agents can be very complex and tracking the behavior requires a disciplined approach. At the same time, equation-based models (EBMs) can be more tractable and allow for at least partial analytical insight. However, inadequate representation of the detailed population structure can lead to spurious results, especially when the epidemic process is beginning and individual variation is critical. In this paper, we demonstrate an approach that combines the two modeling paradigms and introduces a hybrid model that starts as agent-based and switches to equation-based after the number of infected individuals is large enough to support a population-averaged approach. This hybrid model can dramatically save computational times and, more fundamentally, allows for the mathematical analysis of emerging structures generated by the ABM.

1 INTRODUCTION

Choosing an appropriate formalism for an epidemiological model can pose a challenge. The type of description can range from a compartmental model to very detailed ABMs defined by contact networks (Riley, 2007). Some back-to-back comparison of equation-based and agent-based models of dynamics of contagion was done in Rahmandad and Sterman (2007). Equation-based approaches, such as compartmental models, operate on global laws defined by the equations and apply to all members of the compartment. Adding stochasticity does not change the description in principle, but rather utilizes the concept of independent and identically distributed (iid) objects. In a number of situations such as the spread of infectious diseases, especially sexually transmitted diseases, it is important to describe more detailed micro-level transmission processes.

When describing a number of interacting and structured populations, EBMs become more complex and include a large number of equations, which makes them less tractable. Furthermore, in this situation, EBMs require that specific parameters be estimated for each category.

Alternatively, in individual-based models, the rules are defined at the individual level, which allows one to capture local interactions and adaptive behavior (e.g., prevalence elastic behavior). The conclusions about aggregate dynamics are made by integrating the behavior over the ensemble of interacting agents. These models are sometimes called "individual-based models" in epidemiology because the term "agent" has a different meaning (e.g, contagious pathogens).

ABMs thus have the advantage that they explicitly represent epidemic processes at the local level, permitting one to generate macroscopic structures that can then be compared to data.. If one's focus is epidemic behavior at the local level (i.e., small town or village) then aggregated equation-based representations may be too general and hence misleading.

This advantage in capturing local interactions, however, is not without cost: ABMs may impose a heavy computational and parametric burden. Tracking and scheduling a large number of interacting agents leads to serious computational requirements and analytical challenges.

All of this raises a central methodological question: "Do we really need to keep track of local interaction all the way through to the end of the simulations, or is there a natural way to switch to aggregated dynamics after the global structures have emerged?" For example, when more global behavior is of interest, such as country-level disease incidence or prevalence, can the emergent behavior be captured at a higher level of aggregation? Perhaps, after the emergent patterns have been established, they can be summarized in a set of global, equation-based rules such that the population-level outcomes could be adequate to the ones of the ABMs?

In this paper, we present a virtual laboratory to study these questions and provide an example of a mathematical model of pandemic influenza. In Section 2, we briefly introduce past work on the development of mathematical models of influenza. Then, in Section 3, we introduce a concept of a hybrid threshold model. We present demonstration results in Section 4. Finally, conclusions and future work are summarized in Section 5.

2 PAST DEVELOPMENTS IN MATHEMATICAL MODELS OF GLOBAL INFLUENZA

Some of the first models describing the global spread of influenza were developed by Baroyan et al. and then further elaborated by Rvachev and Longini in 1985. The models were based on a system of coupled difference equations. More recent models have expanded the to include more cities around the world and added intervention details (Epstein et al. 2007, Collizza et al. 2006ab, 2007, Cooper et al., 2006). Country-wide models developed by Longini et al. (2005) provide an intermediate step between pure agent-based and patch models. Ferguson et al. (2006) and Parker et al. (2007) have developed pure ABMs for the United States; however, the models incorporate little details about the community structure within the cities. Longini et al, (2004) have developed an ABM for community-based disease transmission where agents travel between community components (school, work, home) according to their age. More detailed models that include community structure were developed by Eubank et al. (2004) under Episims project. Episims has the most detail and describes human activities within a city using respondents' blogs. It includes detailed city structure; traffic patterns; and contact patterns at work, in schools, at homes, at public places, etc.

The difference in model complexity is reflected in the computational resources to run the simulations. For example, an EBM of the Rvachev or Longini type can be easily run on a laptop with 512M of RAM. The same is true for a small ABM with less than 1M agents. However, when the number of agents is large and is on the scale of U.S. size (300M), such as in the models by Parker et al. or Ferguson et al., more computational power is needed, such as a 64-node double processor Linux cluster. Probably the most complexity is involved in the Episims model which requires detailed event scheduling, and thus, has been ran on similar clusters even for an 8M agent city. We are considering a model that would combine the detailed description of the epidemic up to a certain point, and then we would switch to an EBM with much less computational demands.

3 HYBRID THRESHOLD MODELS

3.1 General Concept

The concept of the hybrid threshold model originates from the premise that when the number of active agents is large, the law of large numbers and central limit theorem could be applicable to iid observations. Thus, it should be possible to aggregate the behavior of similar agents and to model their behavior through mean-field approximations.

3.2 Homogeneously Mixed Paradigm in a Single City

We start illustrating our approach with a simple homogeneously mixed epidemic model with no adaptive behavior and no community structure. Although such a simplistic description does not allow ABMs to produce emerging structures, it illustrates the mechanics of switching between agent-based and equation-based formalisms. Each infected agent thus has a chance to meet any other agent and pass an infection to him or her. The rate of disease transmission (λ) is dependent on how many susceptible agents an infected agent meets per unit time (β) and the probability of disease transmission per contact (ρ). Knowing this rate, a deterministic version could be built as it was done in Epstein et al.

3.2.1 Agent-based Model

An individual is considered to be in one of the following four states: (1) Susceptible (can contract the disease given the contact with an infected individual), (2) Exposed (contracted the diseases, but is in a latent state without showing symptoms), (3) Infectious (showing symptoms and capable of infecting others), and (4) Recovered (obtained permanent immunity and cannot infect others). We denote the number of Susceptible, Exposed, Infectious, Recovered, and Total individuals as S,E,I,R,T, respectively.

Because the interaction between Susceptible agents is irrelevant to the epidemic process in the ABM, we track only the contacts of the Exposed and Infectious agents, keeping Susceptible agents inactive. When an infected individual passes infection to a Susceptible individual, a corresponding Susceptible agent is activated. This algorithm allows one to keep the number of active (Exposed and Infectious) agents low at the beginning of the epidemic process and thus decreases the computational burden. As the number of activated agents increases, the amount of required computational resources (e.g., RAM) increases as well.

3.2.2 Equation-based Description

The Equation-based model is based on a compartmental approach where individuals travel between S,E,I,R compartments with rates defined by transition equations. Denoting the time since becoming Infectious as τ, the total number of newly Exposed persons in a city at the start of day $t+1$ is

$$E(0, t+1) = \lambda \frac{S(t)}{T(t)} \sum_{\tau=1}^{\tau \max} I(\tau, t) \qquad (1)$$

where index τ corresponds to τ days of being infectious.

Because of the stochastic nature of the Infectious process the number of new Exposed individuals will be a random number following some distribution. Many approaches were taken to estimate this distribution. For the equation-based case, we included two potential sources of stochasticity in the model: (1) random contact between individuals, and (2) random travel from city to city. Randomness is applied to each of these processes in a way that accounts for the underlying nature of the process involved. We assume that random contacts between pairs of individuals are independent of each other and that the number of new contacts that occur between two times, t and $t + \Delta t$, does not depend on either the number of previous contacts or the time t. Under these assumptions, the number of random contacts between individuals follows a Poisson distribution, where the mean number of Infectious contacts in the case of random contact is equal to the number of Infectious contacts in the deterministic case. At the beginning of an epidemic when say a single individual brings the disease to the city, the epidemic may start or may fade out if the infected individual does not pass the disease to the others. This stochastic effect is especially important at the early stages of an epidemic and can depend on the local contact network of the infected individual. When the number of Infectious individuals is large, the individual effects cancel out and the new numbers of infected can be approximated by the averaged value in (1).

3.2.3 Threshold Switch

When the number of infected individuals reaches certain value we stop the model and evaluate the states in which the agents are found. These values are then used as initial conditions for the EBM and an initial time point is considered the same as the time at which the ABM was stopped.

It is important to understand when to switch from agent-based to equation-based formalism. In our initial experimentation we have explored a number of switching thresholds corresponding to 50, 100, and 200 individuals. Conversely, when the number of infected individuals becomes small we switch back to an agent-based formulation to avoid artifacts that can be caused by an EBM.

We consider the ABM to be a "gold standard" because it has the most micro details. We thus compare average epidemic trajectories produced by stochastic equation-based and an ABMs and determine which threshold value ,if any, makes two average trajectories equivalent.

3.3 Global Epidemic Model (GEM) with Multiple Cities

The GEM considers a network of cities around the world connected by air travel. Population sizes and travel data are obtained from a number of publicly available sources (e.g., Population Division, U.S. Census Bureau 2004 Brinkhoff 2005, Helders 2005). In this model, individuals travel between the cities according to probabilistic rules based on population size and the number of available airline seats. Assuming symmetry in travel, the number of individuals entering and leaving a city is equal on a daily basis. When an individual becomes infected (i.e., shows symptoms) that person is banned from travel with certain probability. Because of the latent period, Exposed individuals are the ones who transmit most of the disease.

3.3.1 Agent-based Description

When no epidemic is present in a city, the city population is in a "dormant" stage with no active agents. When an Exposed or Infectious individual travels to a city, the agent can start an epidemic (i.e., activate new Exposed and Infectious individuals).

3.3.2 Equation-based Description

Our EBM consists of a number (155) of coupled systems of difference equations. Travel is defined as travel flows. Let $A_i(t)$ represent the number of individuals in city i on day t in any of the groups allowed to travel, and let σ_{ij} be the average daily number of travelers from city i to city j. The net change in $A_i(t)$ due to travel to and from city i can be included in a transportation operator Ω:

$$\Omega[A_i(t)] = \sum_{j=1}^{n} \left[A_j(t) \frac{\sigma_{ji}}{T_j(t)} - A_i(t) \frac{\sigma_{ij}}{T_i(t)} \right]$$

$$= \sum_{j=1}^{n} \left[A_j(t) \cdot pT_{ji}(t) - A_i(t) \cdot pT_{ij}(t) \right]$$

The first term in the summation refers to individuals traveling to city i from city j. The second term in the summation refers to individuals traveling from city i to city j. Note that when $j = i$, the terms in the summation

cancel and the net number of travelers from city i to itself is zero.

To prevent the early occurrence of new epidemics in cities due to small fractions of Exposed individuals moving through the transportation network, the definition of Ω is modified slightly for Exposed travelers:

$$\Omega'[E_i(\tau,t)] = \begin{cases} \Omega[E_i(\tau,t)] & \text{if } \Omega[E_i(\tau,t)] \cdot \sum_{\tau=0}^{\tau_1} f(\tau) \geq 1 \\ 0 & \text{if } \Omega[E_i(\tau,t)] \cdot \sum_{\tau=0}^{\tau_1} f(\tau) < 1 \end{cases}$$

This modification ensures that Exposed individuals are allowed to travel to an unexposed city *i* only if the approximate expected total of these individuals is at least one.

3.3.3 Threshold Switch

Because an epidemic in different cities has different dynamics some cities can be at the agent-based state with all agents inactive, other cities can have active agents below the switching threshold, and some cities can have a large epidemic beyond the threshold value and thus follow the equation-based description. For a fixed (and city-dependent) threshold we have developed a set of rules that account for removing a number of traveling Exposed individuals from an equation-based city and placing them in an agent-based city. Similarly, we have addressed the reverse issue when the agents from an agent-based city move to an equation-based city. When both travel and destination cities are of the same type we do not have special means for accounting for the right number of travelers and for keeping the total population constant.

Introduction of quarantine, travel restrictions, and vaccination is adjusted accordingly in each city to match an agent-based or equation-based description.

3.4 Structured Cities

For structured cities we follow Longini (2004) and construct five homogeneously mixed groups in the model:

- Households
- Play groups
- Schools
- Work/office
- Social groups.

Implicitly, age structure is built into the model so that children do not go to work and adults do not go to school. We generated a synthetic population with household, play group, school, and social group compositions. With this city structure, an individual will still go through the four disease states (Susceptible, Exposed, Infectious, and Recovered); however, the degree of contact between different mixing groups will be different, and therefore, affect the disease transmission. We do not consider local transportation and detailed social network. Model performance is similar to the GEM, with individuals traveling between the social groups with the exception that the individuals all move back to their households at night. An additional issue in such a structured model is that the households are fairly small in size and have to be activated together with an Infectious individual. This way, the number of agents is not larger than the number of infected individuals and also includes the members of the households. Nevertheless, the number of active agents is still smaller than the number of agents in the fully ABM where all agents are active.

When an initial infected individual is introduced, we can keep track of an average disease transmission matrix (i.e., track the probabilities of who has infected whom). In the simplest structured city we thus have two age groups (adults and children) and a 2 X 2 matrix of infection rates as in Table 1.

Table 1: Infection Rates

From/To	**Adults**	**Children**
Adults	Paa	Pac
Children	Pca	Pcc

Here the infection rates Paa, Pac, Pca, and Pcc correspond to disease transmission probabilities from adult to adult, adult to child, child to adult, and child to child, respectively. When the epidemic process is at its beginning, the estimates in the matrix will be unstable due to the random effects. When the inputs in the matrix are stabilized, we switch to an age structured EBM that uses this transition matrix.

4 EXAMPLES

In Figure 1 we present examples of stochastic trajectories averaged over 50 replications with different switch values for a global epidemic model. Although both equation-based and agent-based models start at the same time, an agent-based epidemic is developing slower perhaps due to initial stochasticity. When threshold is low, around 100 new infected agents, the model switches to the equation-based setting early and follows the pure equation based trajectory very closely. Switching to equation-based model at a later stage such as the threshold value of 200 makes the model runs closer to the model with the switch at 1000 cases which makes us believe that lifting the threshold further will not lead to principally different results.

Because the threshold value could be different for cities of different sizes and the same switch value doesn't fit all scenarios, we need to develop a rule for dynamic threshold setting. In particular, we monitor a simulation-

based estimator of disease transmission rate (Figure 2). When the number of infected individuals is large enough, the estimator stabilizes around the theoretical value corresponding to an equation-based model. When such stabilization occurs, we consider that time point appropriate for the threshold switch.

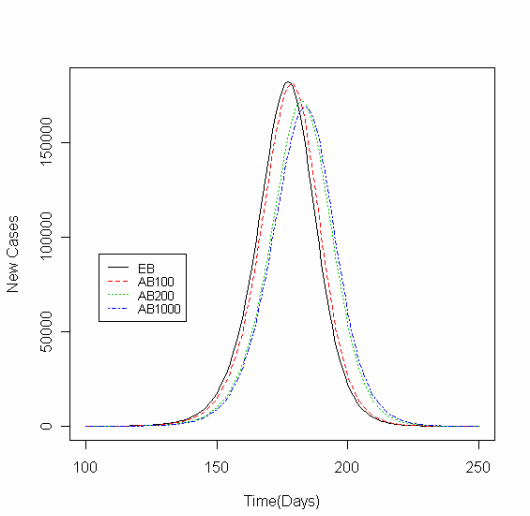

Figure 1. Epidemic curves for Cairo with epidemic start in Hong Kong at day one. The solid line corresponds to the equation-based model, while other lines correspond to threshold model trajectories with threshold switch values corresponding to 100, 200 and 1000 infected individuals.

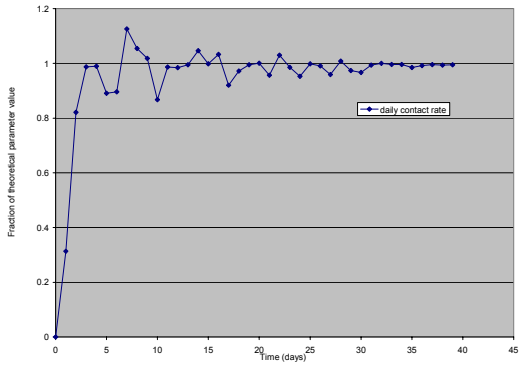

Figure 2. Stabilization of the disease transmission parameter as the number of infected agents is getting larger.

5 DISCUSSION

We have presented a conceptual model and initial examples of a threshold hybrid model that switches between agent-based and equation-based descriptions. The switch allows one to describe the initiation of the epidemic process when the level of uncertainty is high, and allows one to describe efficiently the process of an ongoing epidemic when the level of uncertainty is relatively low. Although the presented examples use a somewhat simplified ABM and do not account for behavior change during an epidemic, the same approach can also be used for a model where all agents are initially active and have adaptive behavior patterns. What is paramount is the definition of the most important transmission subgroups and definition of the transmission matrix. As shown in Heckathorn et al. (1997) for respondent driven sampling, when membership in a group is defined by the contact network, the transmission pattern follows the Markov process. Under the assumption of stationary process, the transmission matrix (i.e., who recruited whom or who infected whom) asymptotically converges to a steady state matrix. Therefore, this matrix can become the basis for a stochastic Markov process operating on the compartmental rather than individual terms. The validity of this approach will be shown in future work.

ACKNOWLEDGMENTS

This work was supported in by the Pilot Studies of Modeling of Infectious Disease Agents Study (MIDAS) cooperative agreement from NIGMS, (1U01 GM070698).

We thank Diane Wagener and R. Joey Morris for valuable discussions.

REFERENCES

Anderson RM, May RM (1991) New York: Oxford University Press. Infectious Diseases of Humans: Dynamics and Control.

Baroyan OV, Mironov GA, Rvachev LA. (1981) An algorithm modeling global epidemics of mutant origin. Programming and Comput Software 6: (5)272–277. English translation from Programmirovanie, 5: 73–79 (1980) (in Russian)

Brinkhoff T (2005) Mato Grosso City Population. Retrieved from http://www.citypopulation.de/Brazil-MatoGrosso.html

Colizza V, Barrat A, Barthélemy M, Vespignani A. (2006a) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci U S A 103: (7)2015–2020.

Colizza V, Barrat A, Barthélemy M, Vespignani A. (2006b) The modeling of global epidemics: Stochastic dynamics and predictability. Bull Math Biol 68: 1893–1921. doi:10.1007/s11538-006-9077-9

Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A. (2007) Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions. PLoS Med 4: (1)e13. doi:10.1371/journal.pmed.0040013

Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. (2006) Delaying the international spread of pandemic influenza.

PLoS Med 3: (6)e212. doi:10.1371/journal.pmed.0030212

Epstein JM, Goedecke DM, Yu F, Morris RJ, Wagener DK, et al (2007) Controlling Pandemic Flu: The Value of International Air Travel Restrictions. PLoS ONE 2(5): e401. doi:10.1371/journal.pone.0000401

Eubank S, Guclu H., Kumar V. S. A., Marathe M.V., Srinivasan A., Toroczkai Z, and Wang N. (2004) Modelling Disease Outbreaks in Realistic Urban Social Networks. Nature, 429: 180–184.

Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437: 209–214.

Ferguson NM, Cummings DAT, Fraser C, Cajka JC, Cooley PC, et al. (2006) Strategies for mitigating an influenza pandemic. Nature 442: 448–452. doi: 10.1038/nature04795

Grais RF, Ellis JH, Glass GE. (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. Eur J Epidemiol 18: 1065–1072.

Heckathorn D.D., (1997) Respondent-driven sampling: A new approach to the study of hidden populations, Social Problems 44 (1997), pp. 174–199

Helders S (2005) World Gazetter.. Retrieved April 20, 2006 from http://www.world-gazetteer.com

Longini IM., Halloran ME., Nizam A. and Yang Y. Containing Pandemic Influenza with Antiviral Agents Am J Epidemiol 2004; 159:623-633.

Longini IM Jr, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, et al. (2005) Containing pandemic influenza at the source. Science 309: 1083–1087.

Mongabay. (2004) World Population Figures Retrieved April 24, 2006 from <http://population.mongabay.com>

Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population Prospects. (2004) World Urbanization Prospects: The 2003 Revision Population Database.. Retrieved April, 2006 from <http://esa.un.org/unup> .

Population Division, U.S. Census Bureau. (2004) Table 1. Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2000 to July 1, 2004 (CBSA-EST2004-01). Retrieved April, 2006 from <http://www.census.gov/population/www/estimates/Estimates%20pages_final.html> .

Rahmandad, H. and J. D. Sterman. (2006). Hetergeneity and network structure in the dynamics of contagion: Comparing agent-based and differential equation models. Available at <http://filebox.vt.edu/users/hazhir/www/papers/Rahmandad-Sterman_070321.pdf>.

Riley S. (2007) Infectious Disease Large-Scale Spatial-Transmission Science 316, 1298; DOI: 10.1126/science.1134695

Rvachev LA, Longini IM (1985) A mathematical model for the global spread of influenza. Math Biosci 75: 3–22.

AUTHOR BIOGRAPHIES

GEORGIY V. BOBASHEV is a Senior Research Statistician at RTI International and a Non-resident fellow at Brookings Institution. He received his PhD in Biomathematics from North Carolina State University. Dr. Bobashev directs and manages statistical and mathematical projects that employ both deterministic and stochastic modeling. His research interests are in modeling and analysis of complex systems with a focus on health-related issues. His e-mail address is <bobashev@rti.org>.

D. MICHAEL GOEDECKE is a Research Statistician at RTI International and a Non-resident fellow at Brookings Institution. He received his PhD in Biomathematics from North Carolina State University. His research interests are focused on the mathematical modeling of biological systems at either the cellular or population levels. His e-mail is <mgoedecke@rti.org>.

FENG YU is a Research Statistician at RTI International. She received her PhD in Chemical Engineering from the University of Tennessee, Knoxville and her MS in Statistics from the University of Nebraska-Lincoln. Feng has been involved in numerous statistical and modeling projects. Her research interests include mathematical modeling in engineering and epidemiology and statistical analysis of complex survey data. Her e-mail address is <fyu@rti.org>.

JOSHUA M. EPSTEIN is Director of the Center on Social and Economic Dynamics and Senior Fellow in Economic Studies at The Brookings Institution, and a member of the External Faculty of The Santa Fe Institute His current projects include modeling of complex social, economic and biological systems using agent-based computational models and nonlinear dynamical systems. Epstein attended Amherst College, and received his Ph.D. from the Massachusetts Institute of Technology. His books include *Generative Social Science: Studies in Agent-Based Computational Modeling* (Princeton University Press, 2006); *Growing Artificial Societies: Social Science From the Bottom Up*, with Robert Axtell (MIT Press/Brookings Institution, 1996); *Nonlinear Dynamics, Mathematical Biology, and Social Science* (the Santa Fe Institute/Addison-Wesley, 1997). His e-mail address is <jepstein@brookings.edu>.