Comparing Compartment and Agent-based Models

Shannon Gallagher

JSM Baltimore, MD

August 2, 2017

Thesis work with:
William F. Eddy (Chair)
Joel Greenhouse
Howard Seltman
Cosma Shalizi
Samuel L. Ventura

Goal: Combine two good models into a better one

Studying infectious disease is important

Compartment vs. Agent-based Models

Compartment models (CMs) describe how individuals evolve over time

Assumptions **(Anderson and May 1992)** :

1. Homogeneity of individuals

Compartment models (CMs) describe how individuals evolve over time

Assumptions **(Anderson and May 1992)** :

1. Homogeneity of individuals
2. Law of mass action
   $I(t+1) \propto I(t)$

Agent-based models (AMs) simulate the spread of disease

Assumptions **(Helbing 2002)**:

1. Heterogeneity of agents

Agent-based models (AMs) simulate the spread of disease

Assumptions (Helbing 2002):

1. Heterogeneity of agents
2. Model adequately reflects reality

CMs and AMs: a side by side comparison

CMs

- Equation-based
- Computationally fast
- Homogeneous individuals
- No individual properties

AMs

- Simulation-based
- Computationally slow
- Heterogeneous individuals
- Individual properties

Combining the two together

(Bobashev 2007, Banos 2015, Wallentin 2017)

- ad hoc approaches
- perspective from non-statisticians

Combining the two together

(Bobashev 2007, Banos 2015, Wallentin 2017)

- ad hoc approaches
- perspective from non-statisticians

Goal: Create a statistically justified hybrid model

Current Work

There are two main avenues of improvement

1. Quantifying how similar CMs and AMs are

2. Speeding up AM run-time

The SIR model: a detailed look

(Kermack and McKendrick 1927)

$$\begin{cases} \frac{dS}{dt} & = -\frac{\beta SI}{N} \\ \frac{dI}{dt} & = \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} & = \gamma I \end{cases}$$

- β – rate of infection
- γ – rate of recovery
- N – total population size

The SIR model: a detailed look

(Kermack and McKendrick 1927)

$$\begin{cases} \frac{\Delta S}{\Delta t} &= -\frac{\beta SI}{N} \\ \frac{\Delta I}{\Delta t} &= \frac{\beta SI}{N} - \gamma I \\ \frac{\Delta R}{\Delta t} &= \gamma I \end{cases}$$

- β – rate of infection
- γ – rate of recovery
- N – total population size

Our stochastic CM approach

$$
\begin{aligned}
\hat{S}(t+1) &= \hat{S}(t) - s_t \\
\hat{R}(t+1) &= \hat{R}(t) + r_t \\
\hat{I}(t+1) &= N - \hat{S}(t+1) - \hat{R}(t+1),
\end{aligned}
$$

with

$$
\begin{aligned}
s_{t+1} &\sim \text{Binomial}\left(\hat{S}(t), \frac{\beta I(t)}{N}\right) \\
r_{t+1} &\sim \text{Binomial}\left(\hat{I}(t), \gamma\right).
\end{aligned}
$$

Our stochastic AM approach

For an agent $x_n(t)$, $n = 1, 2, \dots, N$, the forward operator for $t > 0$ is

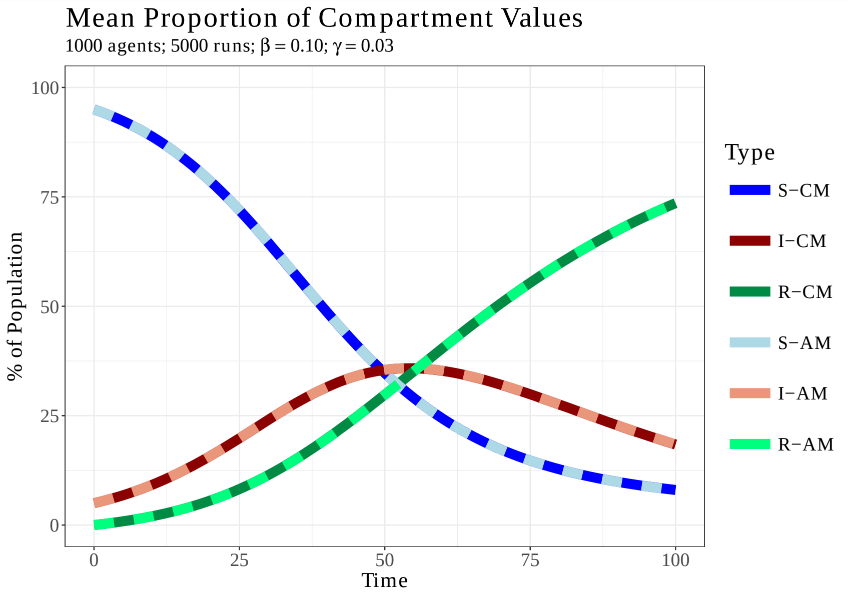
$$x_n(t+1) = \begin{cases} x_n(t) + \text{Bernoulli}\left(\frac{\beta I(t)}{N}\right) & \text{if } x_n(t) = 1 \\ x_n(t) + \text{Bernoulli}(\gamma) & \text{if } x_n(t) = 2 \\ x_n(t) & \text{otherwise} \end{cases} .$$

where $x_n(t) = k$, $k \in \{1, 2, 3\}$ corresponds to state S, I, and R, respectively

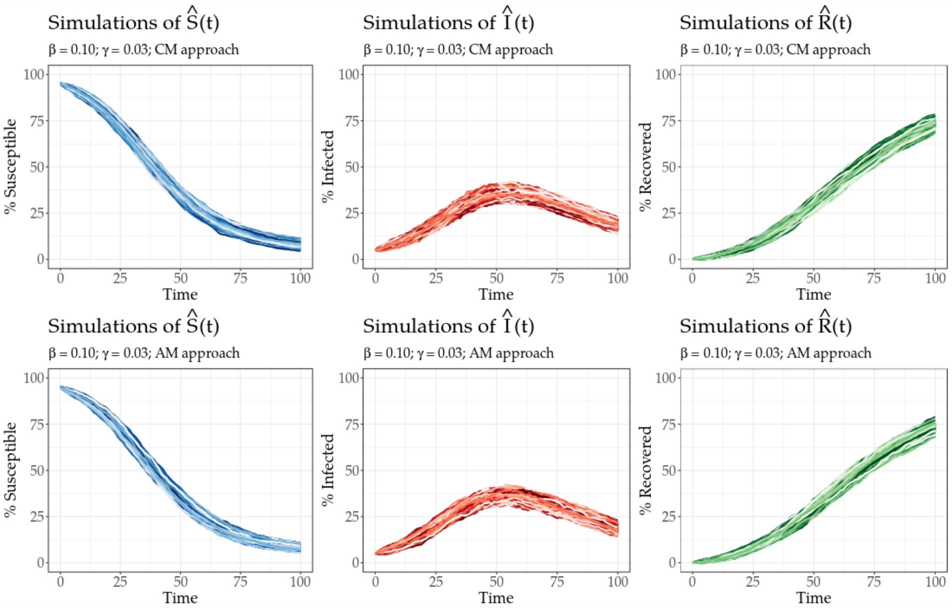
Let the aggregate total in each compartment be

$$\hat{X}_k(t) = \sum_{n=1}^{N} \mathcal{I}\{x_n(t) = k\}$$

The means overlap



The distributions look the same



These approaches are equivalent

Theorem

Let the CM and AM be as previously described. Then for all $t \in \{1, 2, \ldots, T\}$,

$$\begin{aligned} \hat{S}(t) &\overset{d}{=} \hat{X}_S(t) \\ \hat{I}(t) &\overset{d}{=} \hat{X}_I(t) \\ \hat{R}(t) &\overset{d}{=} \hat{X}_R(t). \end{aligned} \tag{1}$$

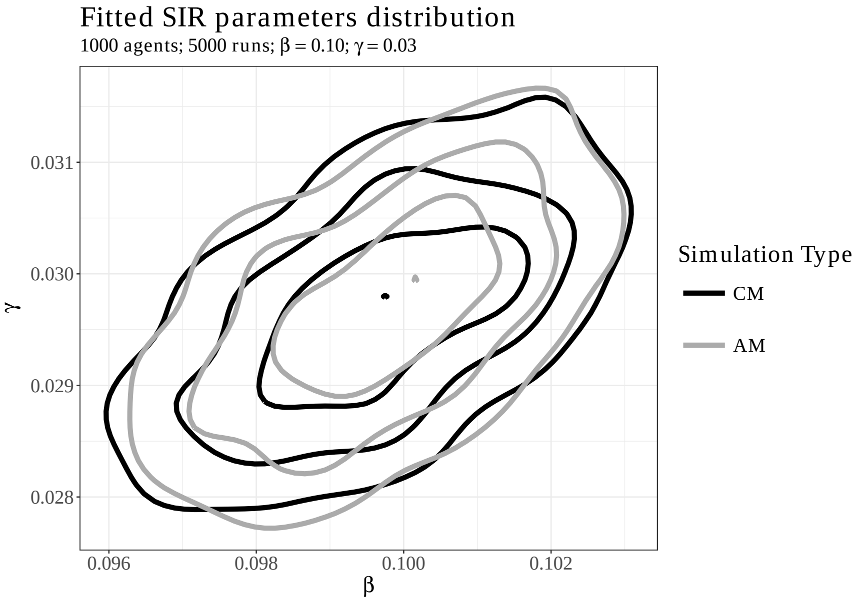
These approaches are equivalent

Theorem

Let the CM and AM be as previously described. Then for all $t \in \{1, 2, \ldots, T\}$,

$$\begin{aligned} \hat{S}(t) &\overset{d}{=} \hat{X}_S(t) \\ \hat{I}(t) &\overset{d}{=} \hat{X}_I(t) \\ \hat{R}(t) &\overset{d}{=} \hat{X}_R(t). \end{aligned} \tag{1}$$

We can compare CM/AM pairs and AM/AM pairs by fitting the underlying model



AMs are appealing because they can be run multiple times

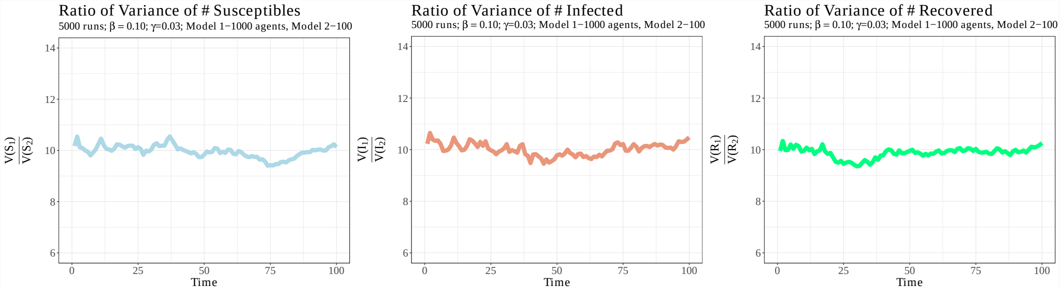
- Simulate an epidemic en masse!
- A run - same initial parameters, different random numbers
- Runs (L) are independent of one another $\implies$ **parallelization**
- Roughly, the variance of compartments $\downarrow$ when N, L $\uparrow$

AMs are appealing because they can be run multiple times

- Simulate an epidemic en masse!
- A run - same initial parameters, different random numbers
- Runs (L) are independent of one another $\implies$ **parallelization**
- Roughly, the variance of compartments $\downarrow$ when N, L $\uparrow$

Goal: Improve computation time without sacrificing statistical details

There is a tradeoff between the number of agents and number of runs



The calculations show that the variance scales

- Note that for a given β and γ, if $\frac{S_1(0)}{N_1} = \frac{S_2(0)}{N_2} \implies \frac{S_1(t)}{N_1} = \frac{S_2(t)}{N_2}$

The calculations show that the variance scales

- Note that for a given β and γ, if $\frac{S_1(0)}{N_1} = \frac{S_2(0)}{N_2} \implies \frac{S_1(t)}{N_1} = \frac{S_2(t)}{N_2}$
- $V\left[\hat{S}(t+1)\right] = S(t)(1-p_t)p_t + (1-p_t)^2 V\left[\hat{S}(t)\right]$

The calculations show that the variance scales

- Note that for a given β and γ, if $\frac{S_1(0)}{N_1} = \frac{S_2(0)}{N_2} \implies \frac{S_1(t)}{N_1} = \frac{S_2(t)}{N_2}$
- $V\left[\hat{S}(t+1)\right] = S(t)(1-p_t)p_t + (1-p_t)^2 V\left[\hat{S}(t)\right]$
- $V[\hat{S}_2(t)] = \frac{N_2}{N_1} V[\hat{S}_1(t)]$

The calculations show that the variance scales

- Note that for a given β and γ, if $\frac{S_1(0)}{N_1} = \frac{S_2(0)}{N_2} \implies \frac{S_1(t)}{N_1} = \frac{S_2(t)}{N_2}$
- $V\left[\hat{S}(t+1)\right] = S(t)(1-p_t)p_t + (1-p_t)^2 V\left[\hat{S}(t)\right]$
- $V[\hat{S}_2(t)] = \frac{N_2}{N_1} V[\hat{S}_1(t)]$

$$
\begin{aligned}
\frac{V\left[\frac{1}{L_1}\sum_{\text{runs } \ell} \frac{\hat{S}_1(t)}{N_1}\right]}{V\left[\frac{1}{L_2}\sum_{\text{runs } \ell} \frac{\hat{S}_2(t)}{N_2}\right]} &= \frac{L_2 N_2^2}{L_1 N_1^2} \cdot \frac{V[\hat{S}_1(t)]}{V[\hat{S}_2(t)]} \\
&= \frac{L_2 N_2}{L_1 N_1}.
\end{aligned}
$$

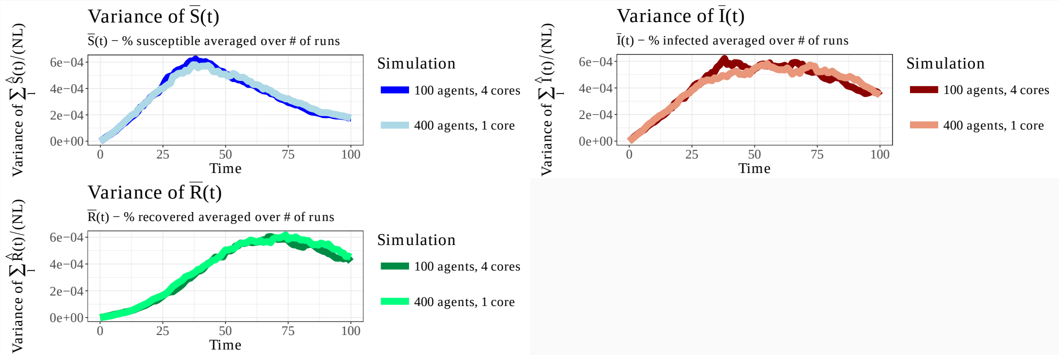
The calculations show that the variance scales

- Note that for a given β and γ, if $\frac{S_1(0)}{N_1} = \frac{S_2(0)}{N_2} \implies \frac{S_1(t)}{N_1} = \frac{S_2(t)}{N_2}$
- $V\left[\hat{S}(t+1)\right] = S(t)(1-p_t)p_t + (1-p_t)^2 V\left[\hat{S}(t)\right]$
- $V[\hat{S}_2(t)] = \frac{N_2}{N_1} V[\hat{S}_1(t)]$

$$\begin{aligned}
\frac{V\left[\frac{1}{L_1}\sum_{\text{runs } \ell} \frac{\hat{S}_1(t)}{N_1}\right]}{V\left[\frac{1}{L_2}\sum_{\text{runs } \ell} \frac{\hat{S}_2(t)}{N_2}\right]} &= \frac{L_2 N_2^2}{L_1 N_1^2} \cdot \frac{V[\hat{S}_1(t)]}{V[\hat{S}_2(t)]} \\
&= \frac{L_2 N_2}{L_1 N_1}.
\end{aligned}$$

We can replace agents with runs!

Through paralellization, we can get a speed-up without losing statistical information



Simulation 1 (100 agents, 4 cores, 100 times): 3:30 minutes

Simulation 2 (400 agents, 1 core, 100 times): 4:05 minutes

Future work

There is more work to be done: short-term

- Implementation of current methods in FRED
  - FRED - an open source, supported, flexible AM
  - Incorporate different levels of homogeneity
    1. Independent agents
    2. Agents go to one other activity (school, work, neighborhood)
    3. Multiple activities
  - Compare CM and AM parameters empirically
- Empirically determine when different regions can be combined

Thank you!

Questions?