Improving performance of large-scale individual-level stochastic disease simulations

Stefan Sellman[a], Kimberly Tsao[b], Michael J. Tildesley[c], Uno Wennergren[a], Colleen T. Webb[b], Tom Lindström[a]

kim.tsao@colostate.edu

Contributed equally

[a]Department of Physics, Chemistry, and Biology, Linköping University, Linköping, Sweden
[b]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
[c]School of Veterinary Medicine and Science, University of Nottingham, United Kingdom

Background

Stochastic disease simulations modeled at the individual level typically require making **numerous pairwise comparisons** to determine if transmission occurs between each infectious and susceptible individual, which in large models can be **extremely slow** and **computing resource-heavy**.

We have adapted a shortcut method from Keeling and Rohani[1] to **substantially reduce computing time** while **maintaining accuracy** of these calculations. This method is illustrated using a simulation of viral spread among livestock premises at the US national scale.

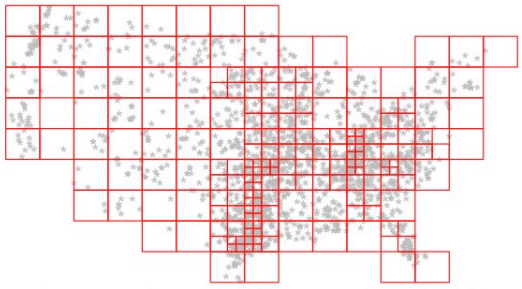
Methods

This method **avoids two time-consuming aspects** of evaluating transmission pairwise:

- Calculating the distance-based probability of exposure for each susceptible individual
- Repeating this same calculation at each time step

- Pre-processing: Lay a **square grid** over fixed point locations – in this example, livestock premises. **Recursively subdivide** cells while the number of farms per cell exceeds a pre-determined threshold, or until the cell is a minimum size[2]. For each cell, **perform and store calculations** that overestimate probabilities of susceptible individuals in that cell becoming exposed.

[1]Modeling Infectious Disease in Humans and Animals, Princeton University Press, 2008, Box 7.3.
[2]Thresholds and sizes for optimal runtimes depend on probabilities of transmission.

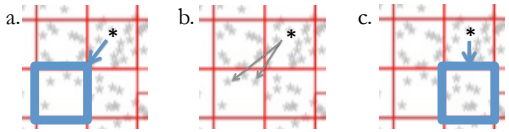


- For each infectious individual:

a. Based on overestimated probabilities, determine the number of susceptible individuals (farms) in a cell that *may* become exposed from this individual.

b. If one or more, **determine which of these individuals actually become exposed**, based on the *true* probabilities of exposure, conditioned on the successful outcome of step a.

c. Otherwise, **skip ahead** to evaluate the next cell.



- Compare results produced by the **shortcut algorithm** at various cell sizes to **pairwise calculations** to ensure parity in infection results. Using each method, generate distributions of:
  - Distances at which infections occurred
  - Total numbers of farms infected
- Calculate and compare the first four moments (mean, variance, skew, and kurtosis) for each distribution.

← Grid cell layout with recursive subdivisions according to density of livestock premises. Grey stars are a subsample of approximately 800000 beef and dairy premises.

Results

Max farms/cell	Method	Mean	Variance	Skewness	Kurtosis
		Distances (m) at which transmission occurred			
n/a	pairwise	933.67	277878.70	207.75	-0.56
n/a	pairwise	933.06	276962.68	210.93	-0.54
500	shortcut	928.10	280650.17	198.96	-0.62
2000	shortcut	929.72	286059.96	207.30	-0.65
		Numbers of farms exposed			
n/a	pairwise	8.14	29.51	0.54	-0.30
n/a	pairwise	7.99	27.02	0.48	-0.43
500	shortcut	8.36	32.23	0.48	-0.53
2000	shortcut	8.07	28.96	0.54	-0.41

↑ Distribution statistics of 1000 replicates of transmission from a single individual at one timestep, using the same parameters and initial conditions for pairwise calculations and the shortcut algorithm.

Under the same conditions and on the same 2.9 GHz laptop processor:

- Runtimes for a single timestep using the shortcut method take **1-10 ms**.
- Runtimes for a single timestep using pairwise comparisons take over **1 hour**.

Conclusions

- This algorithm provides substantial time savings over pairwise calculations, while yielding comparable results of stochastic disease spread.
- This approach is generalizable to other disease-spread simulations with individuals with fixed point locations.

This research has been supported by funding from the US Department of Homeland Security's Science and Technology Directorate, contract number HSHQCDC-13-C-B0028.