A Multi-fidelity Approach for Efficient Network-based Simulation of Disease Outbreaks

Brian M. Adams
Sandia National Laboratories
Optimization and Uncertainty Quantification

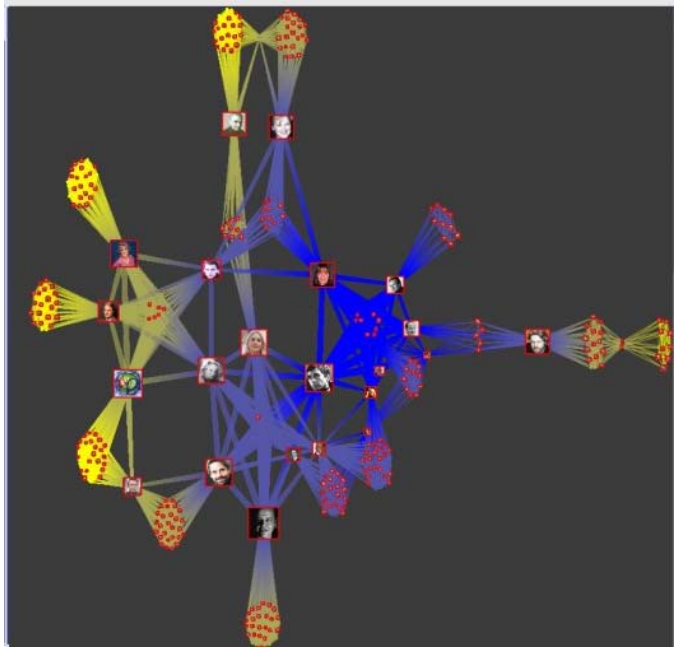
April 26, 2008

Atlantic Coast Symposium on the
Mathematical Sciences in Biology and Biomedicine
Raleigh, NC

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company,
for the United States Department of Energy's National Nuclear Security Administration
under contract DE-AC04-94AL85000.

Outline

Discussion of social contact network-based models for early epoch simulation of natural disease outbreaks and bioterror incidents



http://www.tulip-software.org/

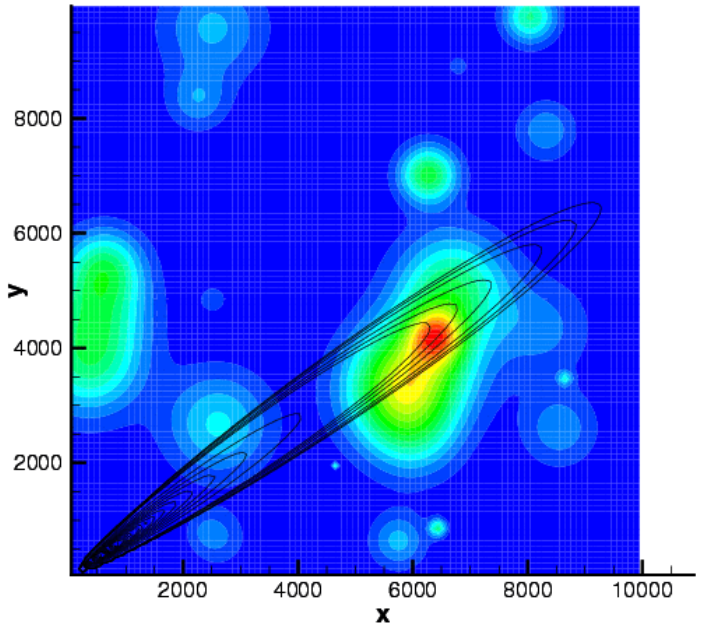
- **GOAL: Outbreak source characterization; analyze potential interventions**
- **Agent-based modeling and suite of social network-based models**
- **Sample model outputs and comparisons**
- **Challenges in model verification and validation *(Why should we believe agent-based models?)***
- ***Acknowledgments: Jaideep Ray (SNL, PI), Karen Devine (SNL), Michael Wolf (UIUC)***

Sandia National Laboratories

Characterize Pathogen Outbreak

- **Pathogen outbreak (natural, aerosolized) undetected by bio surveillance instruments**
- **First intimation: diagnosed symptomatic patients (small numbers)**

- **GOAL: determine initial time, location, and severity of outbreak, e.g., using real time syndromic surveillance data from RODS**
  (7th International Society for Disease Surveillance Conference Dec. 2008, Raleigh)

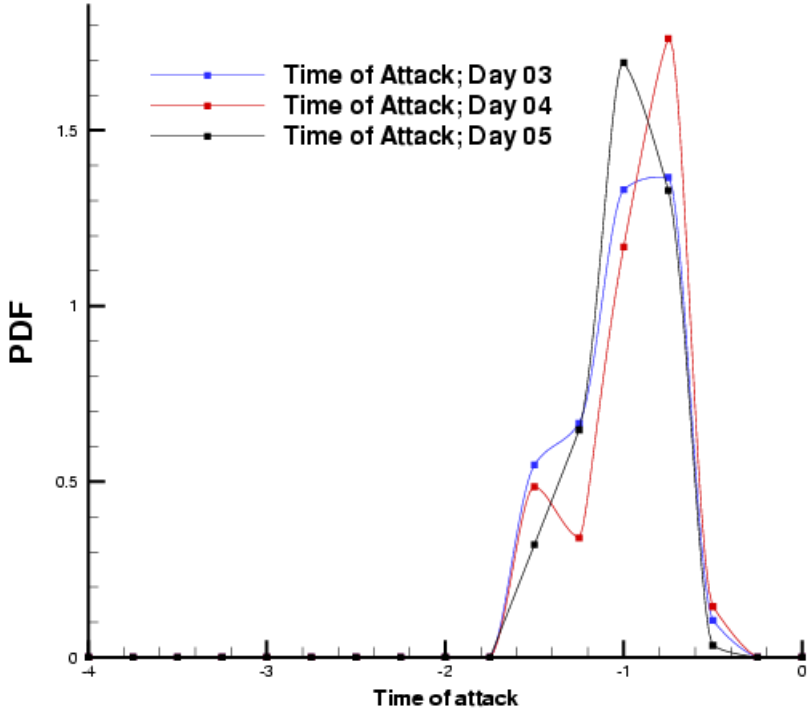


- **Inform public health interventions: rate of spread, where to sanitize, whom to quarantine, where to increase surveillance**
- **Related: provide initial conditions and parameters for real-time simulation-based scenario analysis**

Sandia National Laboratories

Bayesian Characterization

- **Jaideep Ray and Youssef Marzouk: Bayesian methods for source characterization inverse problem: distributions of initial time, number infected, dose received**
- **Challenges: incomplete/limited data, uncertainty in transmission pattern; success for anthrax, Sverdlosk (1979)**
- **Contagious diseases, e.g. smallpox, harder due to long and variable incubation period; determining rate of spread and social network for smallpox epidemic in Abakaliki, Nigeria (1972)**
- **Studying *robustness of inversion process* to assumptions about network structure, number infected, and disease model form error**

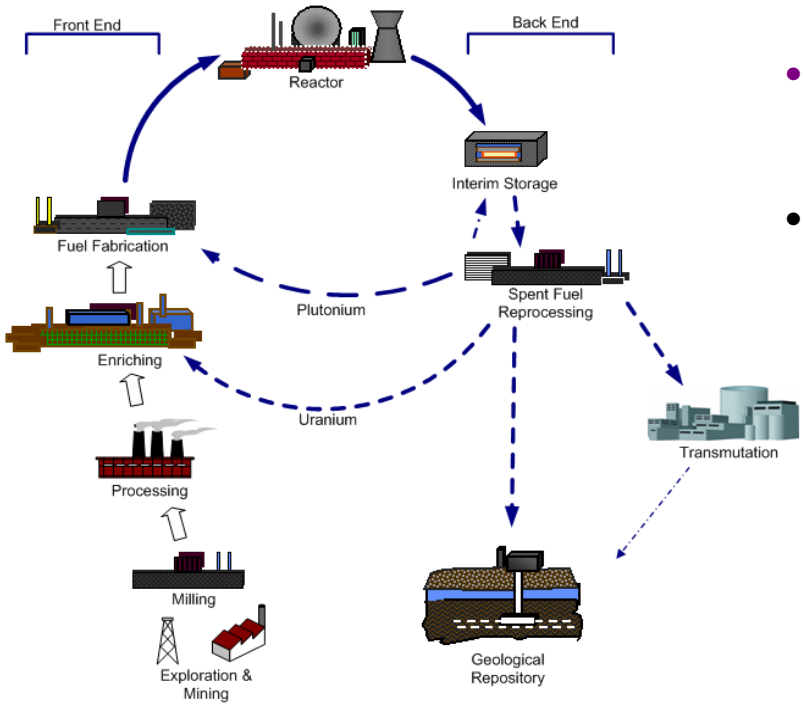


- **Relies on detailed simulations modeling potential chain of transmission, yet sufficient simulation speed to be practical in a Bayesian framework**

Sandia National Laboratories

Agent-based Modeling (ABM)

- ***"An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it [responsive], over time, in pursuit of its own agenda and so as to affect what it senses in the future." ~ Franklin and Greasser, 1996***
- **Agents are typically stochastic with rules for interaction and decision making (could include simple reactionary, psychology, subjectiveness, business motives, learning)**



- ***Ultimately: local learning/adaptation lead to global emergent (complex adaptive) behavior***
- **Intuitive for modeling a variety of systems:**
  - **Social systems and networks (perhaps the only reasonable social models)**
  - **Traffic, transportation systems**
  - **Economy and governments**
  - **Supply chains and logistics**
  - **Systems of systems (e.g., the nuclear fuel cycle, fleet management)**

Sandia National Laboratories

Challenges for ABM

- Agent-based models are intuitive; closely align with natural structures. However, all but simplest transition models are not easily analyzed.
- Lack of information to properly define agent behavior (especially for human decision-making models)
- Stochastic – require ensembles of simulations to characterize outcomes; can be computationally expensive
- Typically not thought of as prescriptive models in the typical engineering sense; no clear way to invert
- Lack of data for agent validation; ultimately the whole system modeled
- Even when data fit possible, what about solution verification, i.e., found the “right” answer to mathematical formulation implemented?

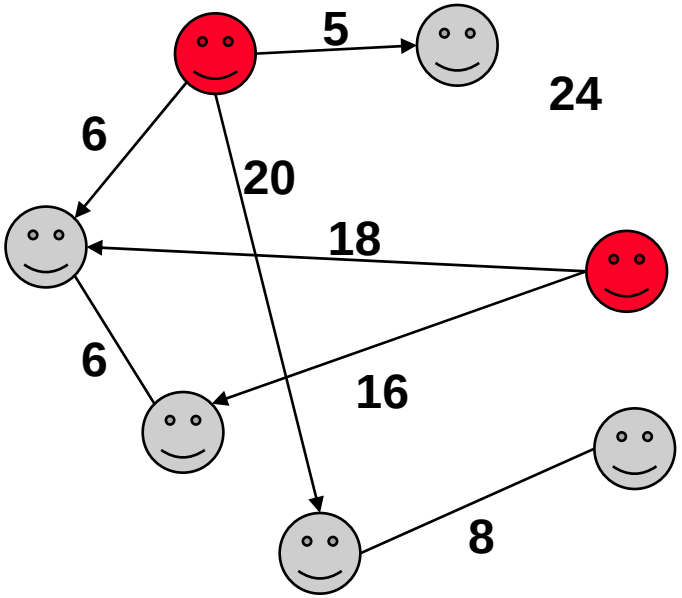
Challenges for ABM

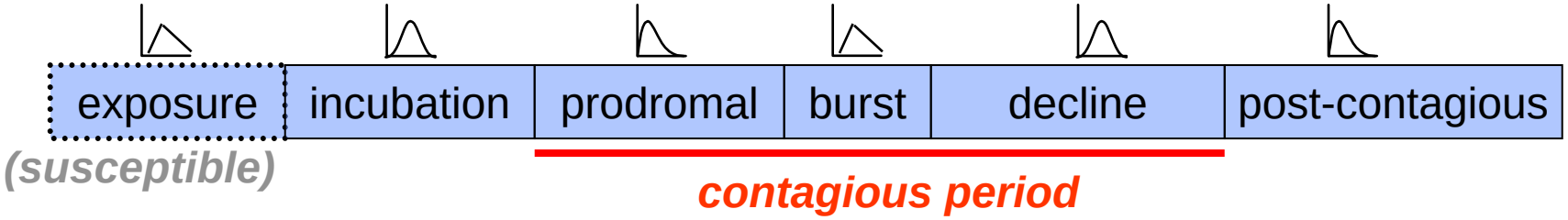
- Agent-based models are intuitive; closely align with natural structures. However, all but simplest transition models are not easily analyzed.

- Despite challenges, ABMs give us transmission chain information and structure necessary to predict other areas of social network where disease will spread.
- Unlike (deterministic) differential equations, potentially valid when well-mixing assumptions invalid.

Simple Contagious Network-based Disease Model

- **Explicitly model each individual**
- **Network edges weighted by (daily) contact time → time-dependent probability of transmission**
- **Model in-host pathogenesis with simple residence time model,**
- **Each period length has a (possibly covariate-dependent) probability distribution across the population**





- ***I refer to the simplest such models as <u>individual-based</u> since, with the exception of choosing when to seek medical care or self-quarantine, disease progression is largely uninfluenced by decision (autonomy).***

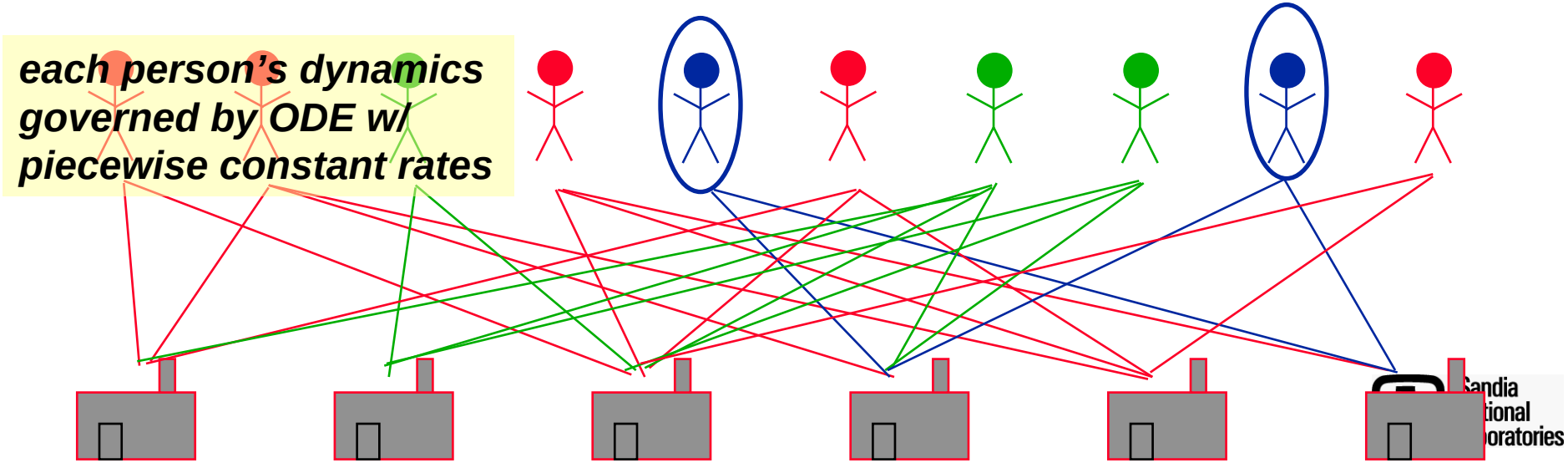
Sandia National Laboratories

Social Contact Disease Spread

- **Recent interest in bioterrorism has renewed interest in network-based disease models, for (selective) example:**
  - ***Burke; Longini; Epstein; others:*** **like prev. slide, parameters census data-based, including assumptions of interaction locations and times**
  - ***Meyers / Pourbohol / Newmann, et al.:*** **Census data-based parameters and structure, but study via bond percolation on random and unstructured graphs**
  - ***Chowell/Castillo-Chavez; surveys: Lloyd, et al.:*** **Study of SIR models and generalizations on networks**
  - ***Eubank/Barrett:*** **EpiSims, transportation data-based, bipartite graph, disease load concept (more complex)**
- **Interest in influenza, plague, smallpox, etc.**

Dynamic Network-based Model

- Dynamic model (EpiSims, Eubank, et al.) based on bipartite, time varying graph: people move among locations throughout a day; shed or absorb virus
- Each person or location has a disease "load"; when exceeds a threshold, progress through infection stages
- EpiSims is Transit network-based; our framework accepts any model for person-person or person-location contact over time (traffic, social networks, demography-based groupings, etc.)



Our Implementation

Early Epoch Disease Model

Goal: Model capable of simulating on dynamic or static contact networks, with load- or probability-based transmission.

- **Flexible to simultaneously track various diseases (e.g., smallpox, plague, influenza, anthrax, bacteria)**
- **Detailed to capture individual agent behavior – crucial for determining how the first few index cases came to be infected**
- **Yet computationally efficient (fast) – to be useful in response planning, needs to determine sources or analyze scenarios in minutes to hours**
  (leverage reduced-order modeling, scalable parallel computing)
- **Enable ensemble simulation and worst-case analysis (mean behavior might not be crucial)**

Sample Outbreak Simulation

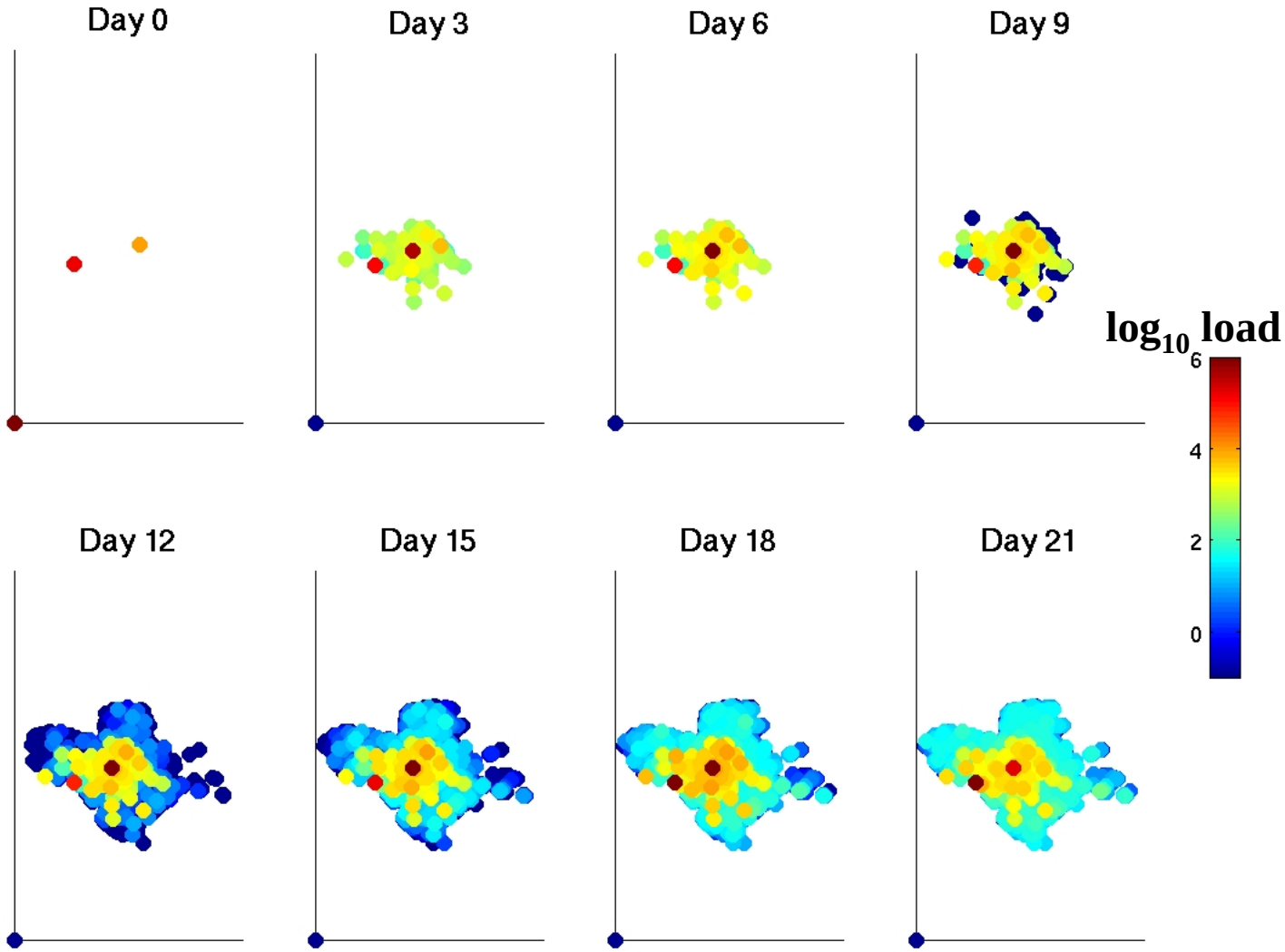
Simulation structure

- **~1.6 million total people, ~200,000 locations**
- **21 day simulation, including ~6000 events per day describing movement of people throughout the city**
- **Uses EpiSims (Eubank, et al.) urban network data, based on transportation data (avail. from LANL/Virginia Tech NDSSL)**

Simulation results for disease spread based on

- ***1599 collocated people receiving 100 plaque-forming units (PFUs) smallpox each,***
- ***312 receiving 30 PFUs at another location, and***
- ***one additional location contaminated with 1000 PFUs.***

Load-based Smallpox Simulation, Dynamic Network: Spatial Spread

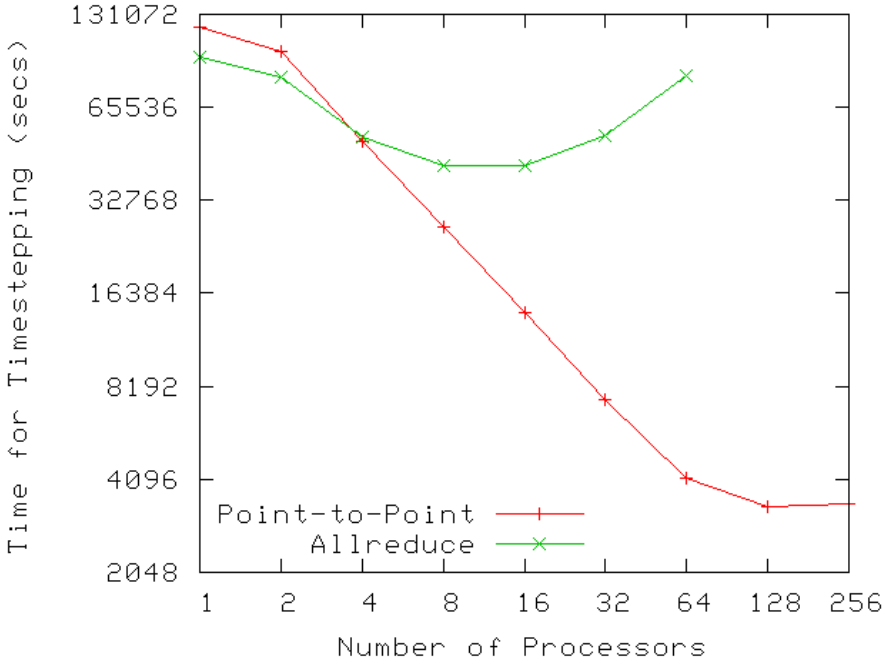


simulated spread of smallpox in an urban setting over 21 days (intensity is sum of person loads at each location)

Sandia National Laboratories

Load-based Smallpox Simulation, Dynamic Network: Parallel Scalability

- **Improvement: partition locations and people; communicate point-to-point between procs needing specific information *(SNL's Zoltan)***
- **Scales well to 64 CPUs**
- **Small (1.6M people, 200K locs) dataset likely preventing further scalability**
- ***Enables more rapid scenario analysis***



Sandia National Laboratories

Potential Model Reductions

Assess feasibility of inverse problem with reduced order (surrogate, meta-) models

- **Approximate dynamic network with static (project in time)**
- **Replace computationally-intensive load-based models with simpler probability of transmission models (age and/or time-dependent)**
- ***Use graph partitioning / clustering to determine key entities and macromodels (aggregate people with similar contacts/locations); sample people or locations***

Use representatives to simulate propagation more quickly while retaining disease characteristics of the full simulation.

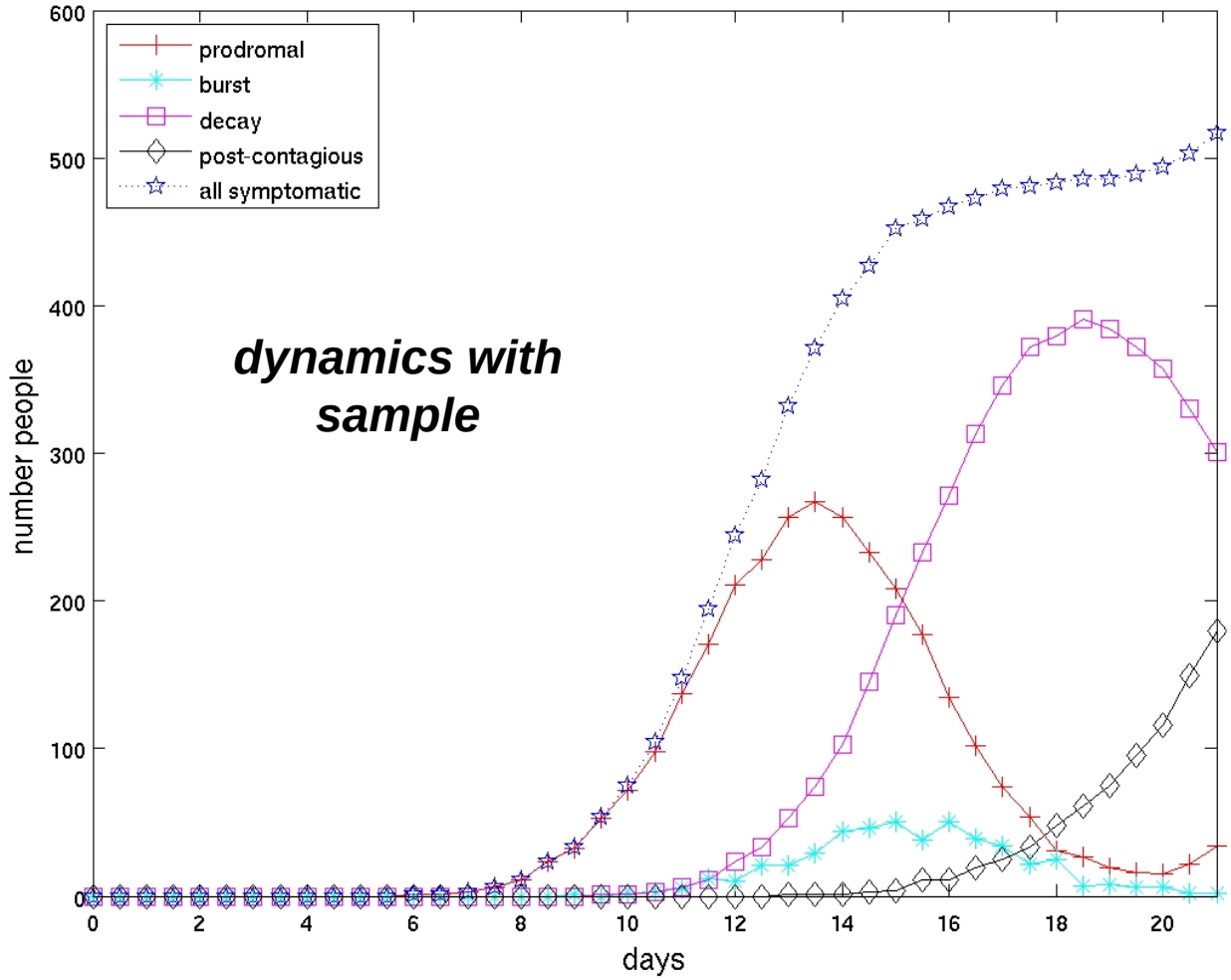
Two Sampling Approaches

- **Person-to-location clustering (K. Devine, Zoltan framework)**
  - **Project visits from 24-hour period to single interaction matrix**
  - **Can represent visits to contaminated locations**
  - **Cannot represent collocation of people with respect to time**
  - **Data size is *M*x*N* for *M* locations and *N* people**

- **Person-to-person graph clustering (M. Wolf)**
  - **Connect each person to other people who are in the same location at same time**
  - **Can represent co-location of people with respect to time**
  - **Cannot represent visits to contaminated locations**
  - **Extremely large data set to cluster – *N*x*N* for *N* people**

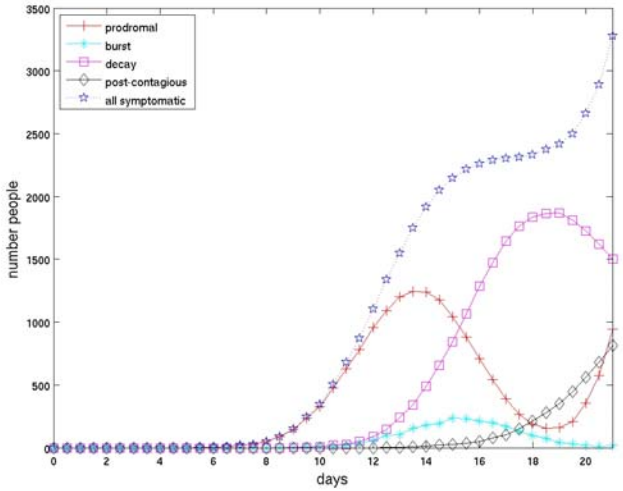
Person to Person Clustered
(238,371 person sample)

- **Characteristics similar to full 1.6 M person simulation**
- **Compute time approx. 10 hrs. on 1 CPU (vs. 40+ hrs. for full simulation)**



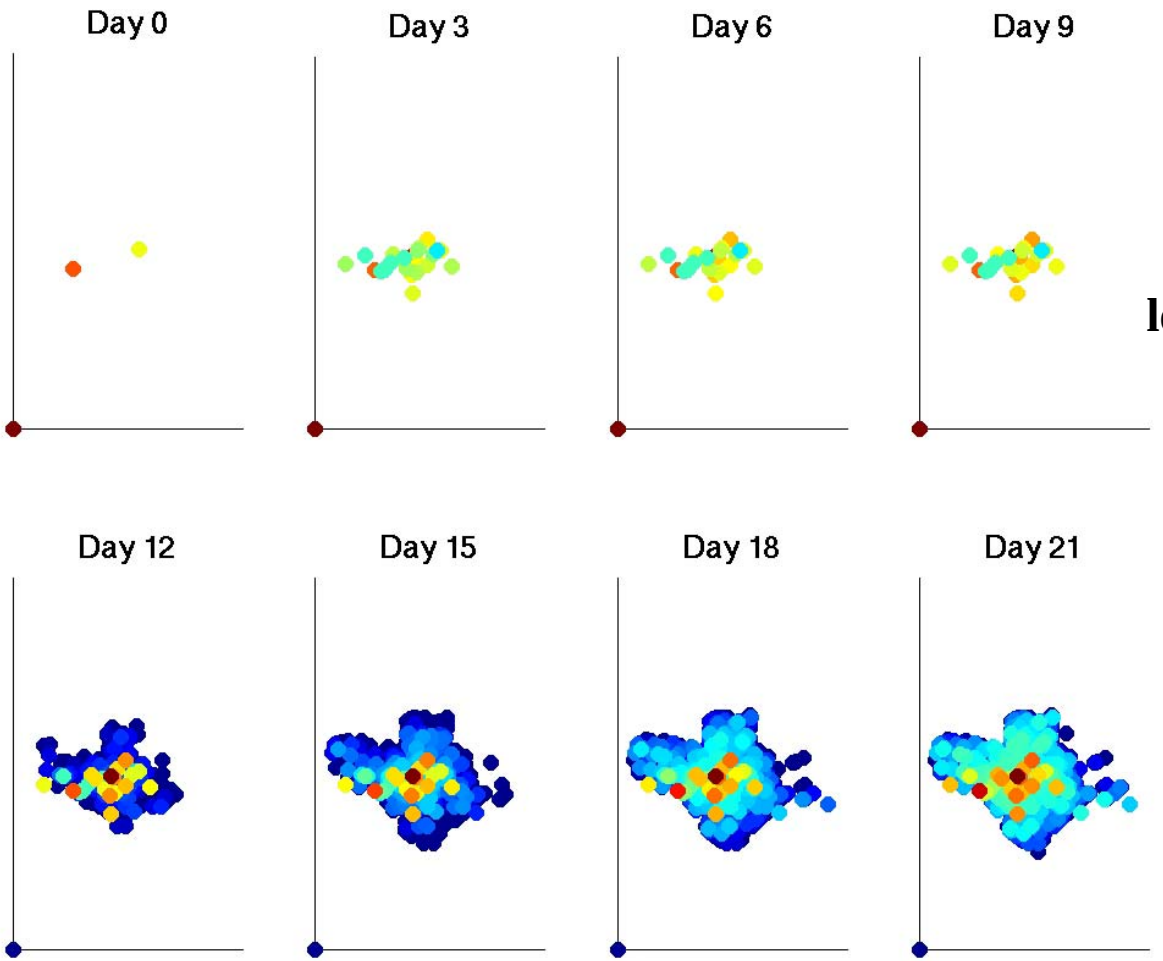
dynamics with sample

full 1.6 M people

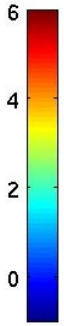


Person to Person Clustered
(238,371 person sample)

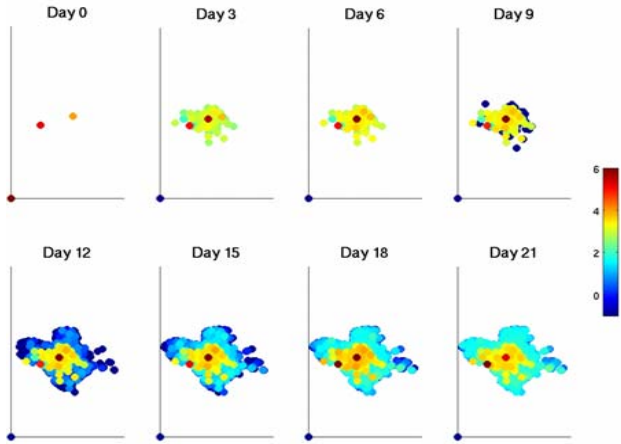
- **Different spatial prediction than full 1.6 M person simulation.**
- **Sampling likely under represents effects of outliers.**



$\log_{10}$ load



full 1.6 M people

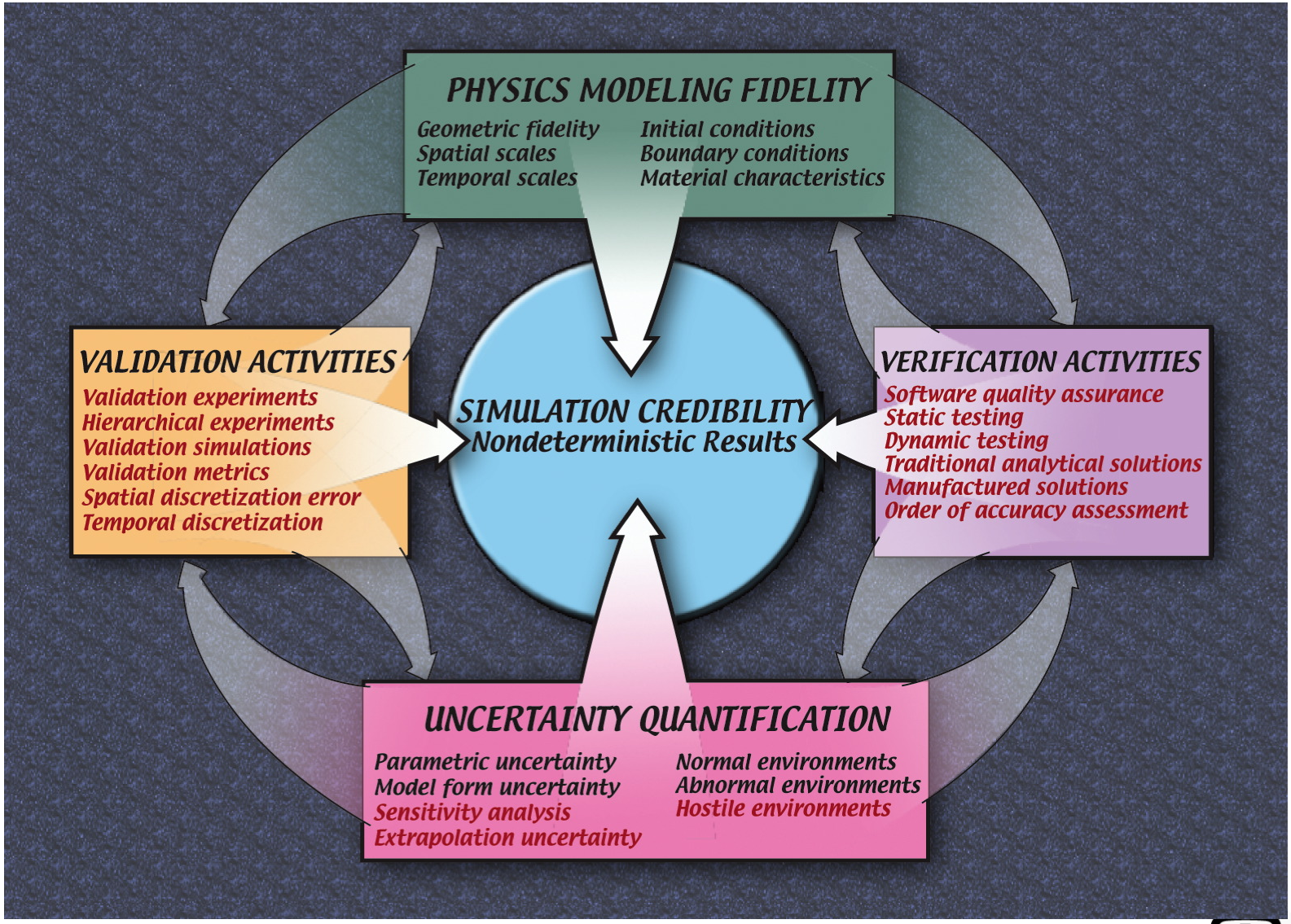


Credible Simulation

- **Ultimate purpose of modeling and simulation is insight, prediction, and decision-making → *need credibility for intended application***

- **Don't want to be criticized for needless complexity: "If they used Occam's razor to design this, it must be growing dull."**

Credible Simulation



Slide credit: Bill Oberkampf

Sandia National Laboratories

Verification & Validation

- **Verification: "Are we solving the equations correctly?"**
  - **mathematics/computer science issue: Is our mathematical formulation and software implementation of the physics model correct?**
  - **code verification (software correctness);**
    solution verification (e.g., exhibits proper order of convergence)

- **Validation – "Are we solving the right equations?"**
  - **a disciplinary science issue: is the science (physics, biology, etc.) model sufficient *for the intended application?* Involves *data and metrics*.**

Related concepts:

- **Sensitivity Analysis (SA): both local and global**
  - **How do code outputs vary with respect to changes in code inputs?**

- **Uncertainty Quantification (UQ):**
  - **What are the probability distributions on code outputs, given the probability distributions on my code inputs? Unknown input distributions?**

- **Quantification of margins and uncertainties (QMU):**
  - **How "close" are my code output predictions (incl. UQ) to the system's required performance level?**

Sandia National Laboratories

V&V Challenges for ABM

- **Code verification: largely (though not entirely) facilitated by good software development practices**

- **Solution Verification: No clear analog to DE order of convergence (with DEs, start with an analytic function, differentiate, and examine residuals from the numerical scheme)**

- **Stochastic simulations – can only examine stats or distributions; need appropriate metrics for ensembles/distributions**

- **Apply hierarchical validation techniques? Need data to validate component and connection models**

- ***Ultimately: give best estimate plus uncertainty, especially when model details not well known***

Sandia National Laboratories

How to perform V&V for ABM

- ***Rudimentary tests might nip at solution verification***
  For example: test network models by creating carefully structured network where transmission and incubation rates are correctly modeled by ODE model: cases where analytic or more well-established solutions apply

- ***Can also test*** **static network structure with simplified transmission models and compare to controlled population (perhaps study rumor spreading or other information dissemination in a controlled setting)**

- ***But don't exercise full intention of model (stochastic, dynamic contacts, emergent behavior, etc.)***

- ***Social networks apply far beyond disease modeling; developing V&Ved predictive capability is crucial.***

Conclusion

- **In appropriate situations, individual-based models are intuitive and useful**
- **Substantial challenges arise in situations where credibility is paramount**
- ***Need a rigorous framework for V&V of such models***
- **Or…**

One way to simplify stochastic simulation?

Thank you for your attention!

briadam@sandia.gov
http://www.sandia.gov/~briadam

Sandia National Laboratories

Abstract

Invited lecture to be presented at
Atlantic Coast Symposium on the Mathematical Sciences in Biology and Biomedicine
North Carolina State University
April 24--26, 2008
Holiday Inn Brownstone Hotel and Conference Center
Raleigh, NC

A Multi-fidelity Approach for Efficient Network-based Simulation of Disease Outbreaks

Presenter: Brian M. Adams (Sandia National Laboratories)
Collaborators: Karen D. Devine and Jaideep Ray (SNL); Michael M. Wolf (UIUC)

Effective public health interventions during a disease outbreak depend on rapid characterization of the initial outbreak and pathways for pathogen spread. Epidemiologically-based modeling and simulation can characterize both and enable practitioners to test intervention strategies. While compartmental differential equation models are often used to represent bulk epidemic properties, they are unsuitable for early time simulations (first few days) when a small number of people are infected (and even fewer symptomatic), nor are they capable of representing spatial disease spread.

In this talk I will present social contact network-based models for early epoch simulation of disease outbreaks and bioterror incidents. They generate detailed predictions of disease spread, including tracking symptomatic individuals, for use in a Bayesian inverse problem context to determine the initial location, time, and size of an outbreak. Our implementation allows for high-fidelity predictive simulations, made efficient though scalable parallelism, as well as reduced-order simulations through population and/or location sampling, and static (time-projected) graph-based simulation, which are practical for inverse problem solution. Efficiently inverting data on early symptomatic presenters characterizes the original source, and the forward simulation can then suggest others who may be at risk.

Sandia National Laboratories