Carnegie Mellon

BioWar: A Scalable Multi-Agent Model of Weaponized Biological Attacks with a High-Degree of Realism

BioWar Group

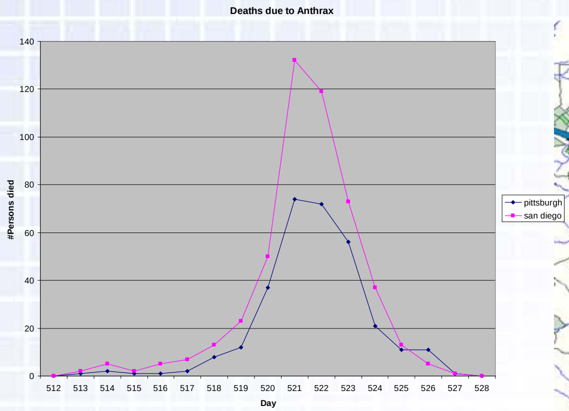
July 14, 2005

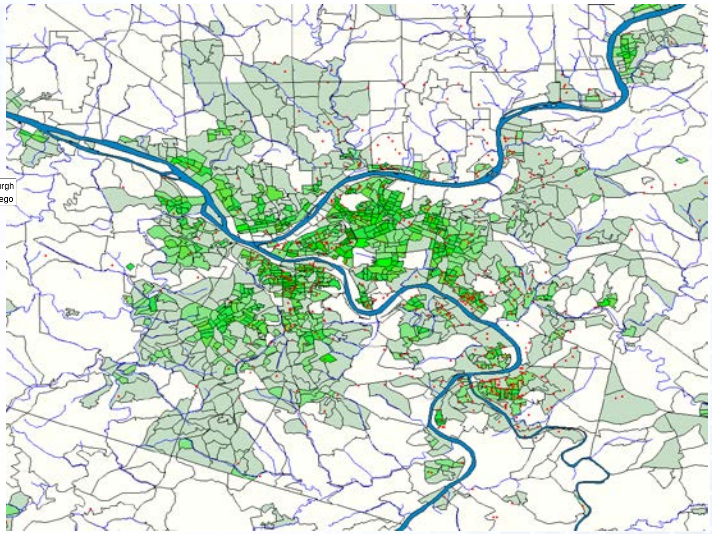
Center for Computational Analysis of Social and Organizational Systems

http://www.casos.cs.cmu.edu/

BioWar – Conceptualization

City Scale Multi-agent Network Model of Weaponized Attacks





Display

Geographic

Chart

Locations
proportional to population

People
10%

Diseases
60 background
2 weaponized

Time
2 year

Attack Profile
-Anthrax medium
- Smallpox medium

Comparison

What Output to Save
- Standard Challenge

Alert Status
- none

City Profile
- Pittsburgh
- San Diego

BioWar Project Team – July 2005

- BioWar – Project Director and PI Kathleen M. Carley, ISRI, Ph.D.
- Elizabeth Casman – CMU, EPP, research scientist

- Graduate Student
  - Alex Yahja, ISRI
- Programming Staff
  - Natasha Kamneva, ISRI, PhD.
  - Neal Altman, ISRI, MSCS.
  - Eric Malloy, ISRI

- Previous Support
  - Jack Chang, PSC, PhD.
  - Li-Chiou Chen – CMU, ISRI, research scientist
  - David Deerfield – PSC, Director of Biomedical Initiative, Ph.D.
  - Douglas Fridsma – University of Pittsburgh, BMSI, Ph.D. M.D.
  - Boris Kaminsky, ISRI, PhD.
  - Démian Nave, PSC, MSCS.
  - Tiffany Tummino, grad student

Objective

- Automated tools for evaluation of response policies, data efficacy, attack severity, and detection tools relating to weaponized biological attacks
- Systematically and automatically reason about:
  - The rate and spread of disease with high degree of realism
  - Early presentation of diseases
  - Potential media and inoculation campaigns (cost, benefit, effectiveness)
  - Other “what-if” occurrence, early detection, and response scenarios
- Policy design with respect to BioWar response
  - What alert-level is appropriate
  - What to do given an alert level, patterns of outbreaks
  - Cost-effectiveness of a policy

Approach

- Realistic and Scalable Multi-Agent Network Model
- Hybrid of many models: agent, network, geographic, weather, disease, diagnosis, etc.
- Utilizes real data streams
  - Census (population and economic)
  - School district, weather, geography, time-budget …
  - Sub-model of military bases
- Multiple outputs
  - OTC purchases, Dr and ER visits, web and phone calls, absenteeism …
  - Outputs at the population and sentinel group level (emergency responders, health care personnel)
- Input and Output stream validation
  - Indicator – input 89.2%; output 63.6%
- 6 Cities (MSAs) have been modeled
  - Hampton Roads, Norfolk, Pittsburgh, San Diego, San Francisco, Washington D.C.

Features

- Input
  - Census data
  - School district data
  - Worksite and entertainment locations & size
  - Hospitals and clinics locations & size
  - Social Network characteristics
  - IT communication procedures
  - Wind characteristics
  - Spatial layout
  - Disease models
    - Influenza, small pox, anthrax, …
- Output
  - Over the counter drug sales
  - Insurance claim reports (Dr. visits)
  - Emergency room reports
  - Absenteeism (school and work)
  - Web access and medical phone calls

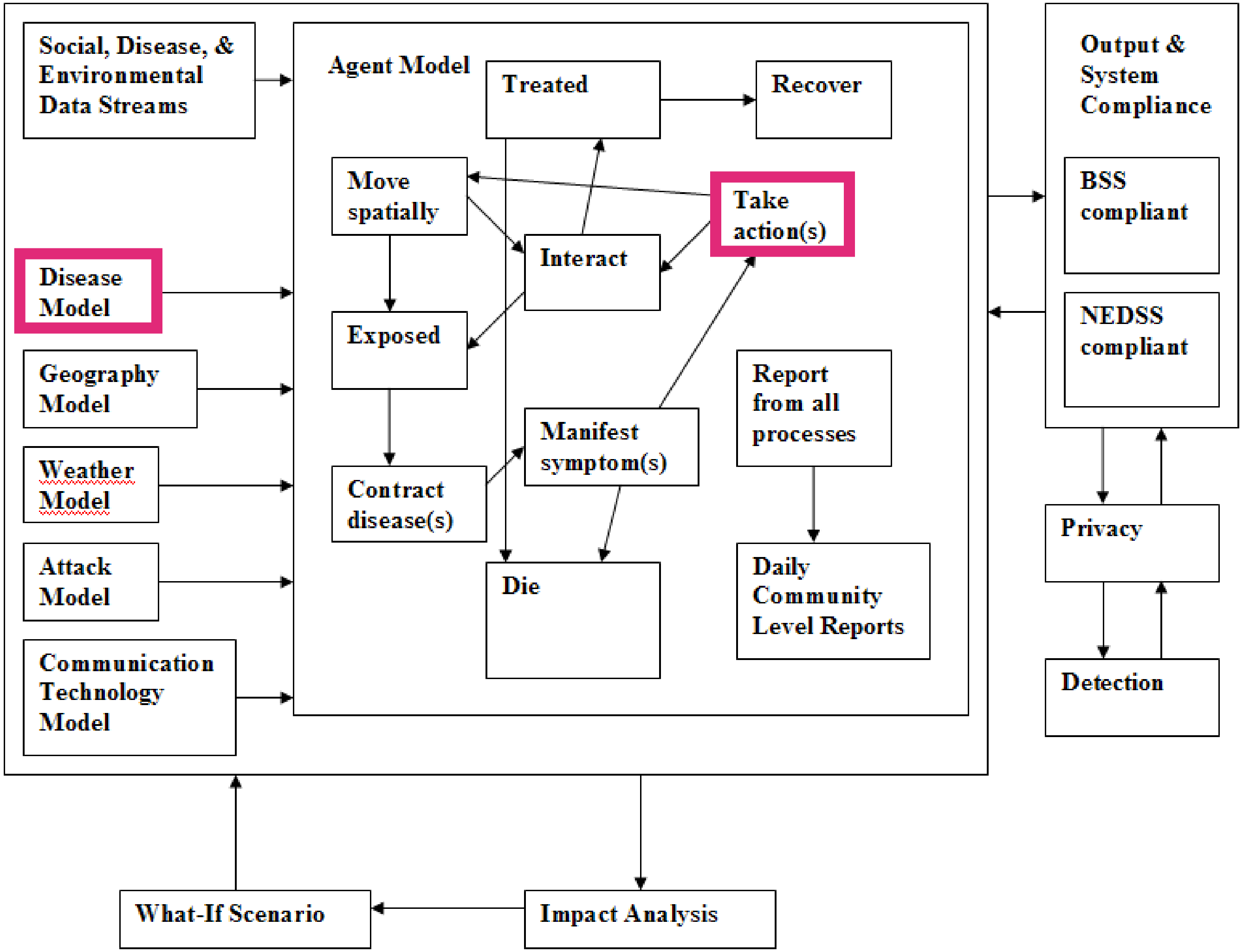
Input Sources

Origin	*Source*	*Description*
USGS	GNIS Database	Hospital, park locations
Census	Summary File 1	Demographics (population, race, age, sex)
	Economic Census	Work, medical, recreation location counts
	Geometry	Cartographic boundaries (region geometry)
NCES	CCD Database	School demographics, locations
	Publications	Student absenteeism statistics
GSS	GSS	Social network characteristics
EPA	www.epa.gov/scram001	Climate, wind data
Internist 1	QMR vocabulary QMR evoking strengths	Disease symptoms, diagnosis model
CDC	NCHS Surveys	Medical visit, mortality & morbidity statistics
CDC	Web sites	Disease timing, symptoms

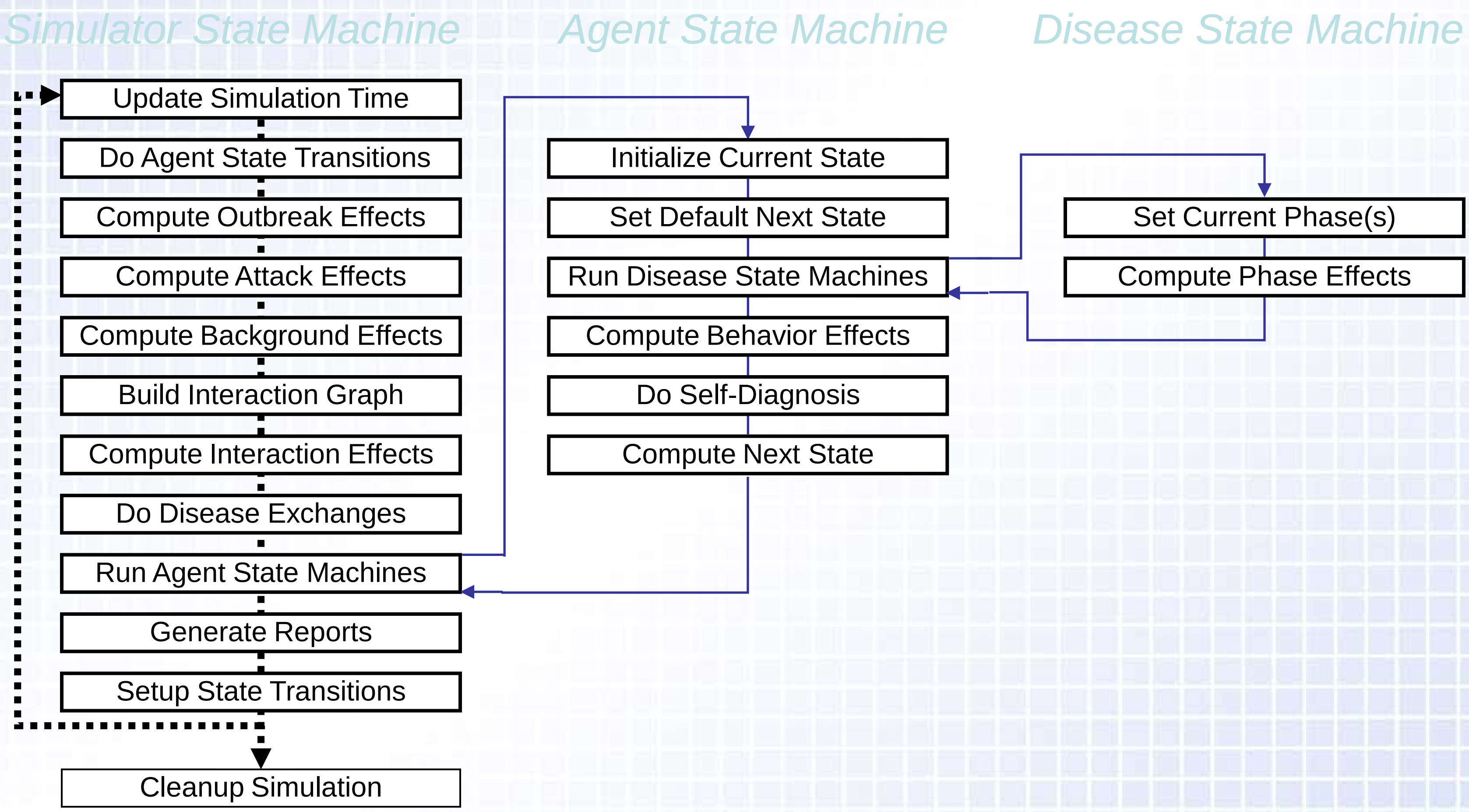
Why an Agent-based Simulation of Disease?

- General-purpose approach to eliminating “homogeneous mixing” assumption
  - Social networks provide mechanism for population mixing on arbitrary number of parameters
- Stochastic nature of the model can generate “unpredictable outcomes” with interaction effects
- Allows modeling at multiple levels (dispersion, response, disease, diagnosis, etc) since *agent* is common across all of these models
- Particularly useful for contagious diseases (SARS, smallpox)
- Particularly useful for examining detection potential of DNA hybridization techniques
- Note: we will give examples of two attack diseases: anthrax and smallpox

BioWar System Diagram



State Machine



CarnegieMellon

Anthrax Disease Model

- The dosage inhaled by the agent is calculated using the following equation:

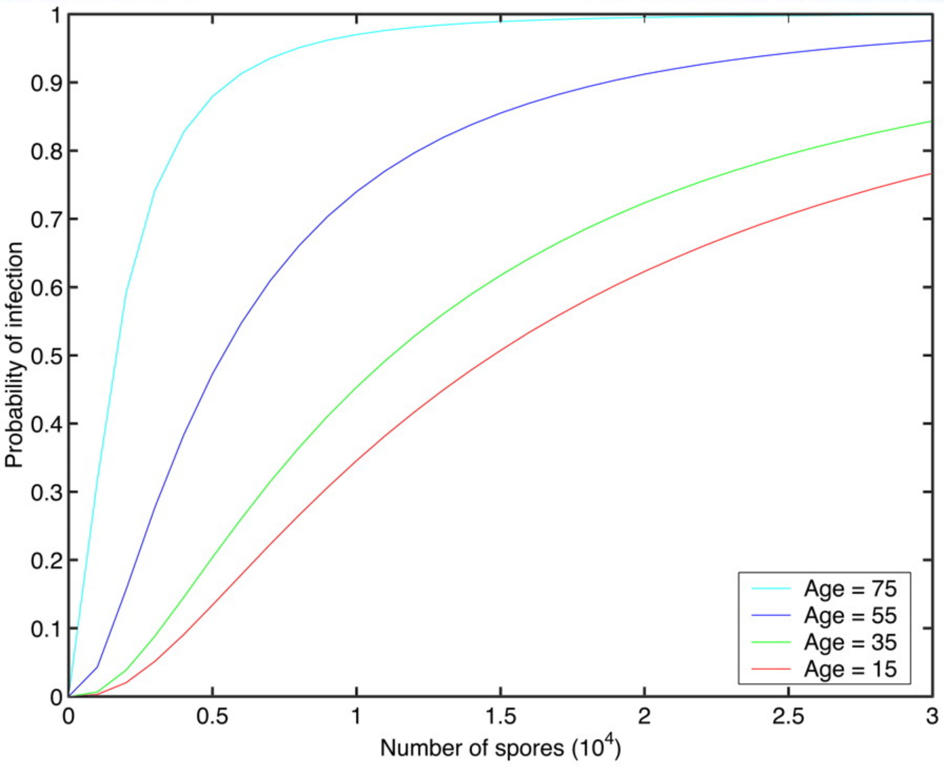
$$Dose = [QB][\pi u \sigma_y \sigma_z]^{-1} exp[-(1/2)(y/\sigma_y)^2] exp[-(1/2)(H/\sigma_z)^2]$$

where *Q* is the source strength (e.g., number of anthrax spores); *B* is breathing rate (usually for light work *B = 5 *10-4 m3/sec*); *u* is wind speed in *m/sec*; σ_y and σ_z are dispersion parameters that are functions of downwind distance *x*; and *H* is height of the release in meters.

$\sigma_y = 0.08x/sqrt(1+0.0001x)$ $\sigma_z = 0.06x/sqrt(1+0.0015x)$

where *x* is the distance downstream of the release point.

- Gaussian Puff model of wind dispersion
- Dose-age response relationship, shown on the side graph
- Lognormal distribution of duration of stages for anthrax (μ for incubation stage = 2.4 days, μ for prodromal stage = 0.85 days, μ for fulminant stage = 0.34 days)



Dose-Age Response Equation

Probability of infection by age:

P[n](S) = b[n] (exp(S/a[n])-1)/(1+b[n](exp(S/a[n]) – 1)),

n=1,2,3,4 (four age categories)

where the two parameters a[n] and b[n] are determined by the infectious doses that produce infections in 50% (ID_{50}) and 10% (ID_{10}) of exposed persons.

Age bracket (years)	ID_{50} infectious dose	ID_{10} infectious dose
<25	15,000	4,500
25-44	10,000	3,000
45-65	6,000	1,800
>65	1,500	450

Symptom-based Behavior

- People who contract an anthrax infection and display fevers/chills may consider their symptoms to stem from influenza/cold, and not significantly alter their behavior. However, if they began having shortness of breath, chest pains, or other symptoms suggestive of a serious problem, they would likely stay home from work, go to doctor, or go to an emergency room.
- A set of symptom severity thresholds guides an agent's decision to visit a medical facility. The thresholds are limits of the sum of the severities of observable symptoms over all diseases infecting an agent, signifying a behavior change:
  - Low severity - no effect
  - Mild severity - go to the pharmacy
  - High severity - go to the doctor
  - Extreme severity - go to the emergency department
- If alerted, individuals will lower their threshold to seek more advanced care.

Validated Features

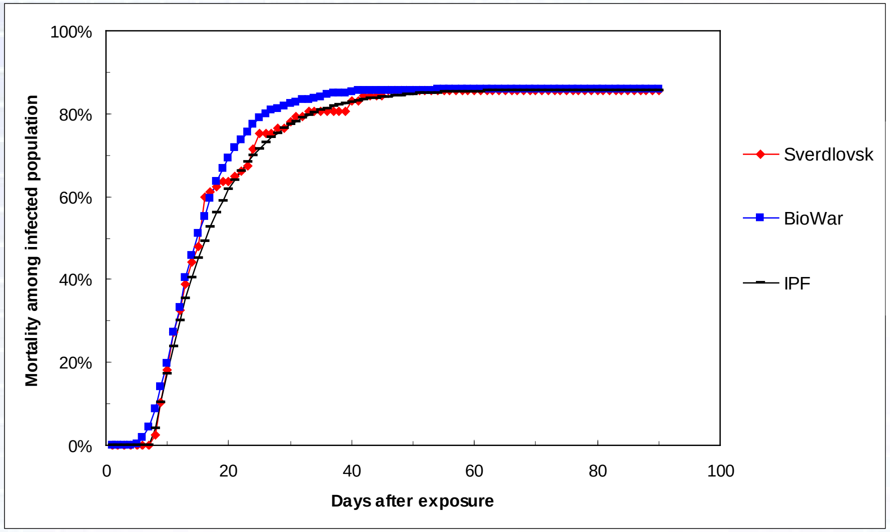
- Social Networks
- Weather (Wind)
- Dispersion
- Anthrax attack & disease model (“docking” or computational model alignment with IPF, Incubation-Prodromal-Fulminant – a revised SIR -- Model)
- Smallpox attack & disease model, docking with SIR
- School absence
- Work absence
- Doctor visit
- ER visit

Validation over Time (C5=last “Challenge”, C1..4=previous “Challenges”)

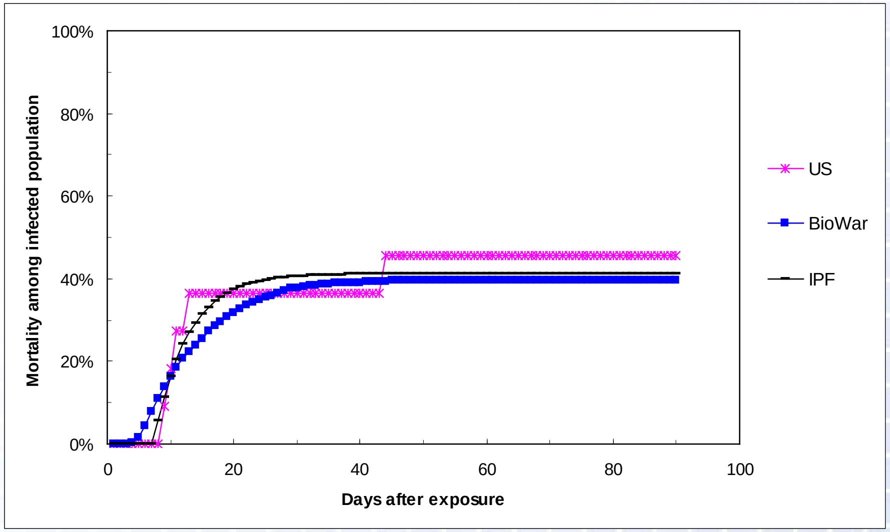
Type	C1	C2	C3	C4	C5
Docking: Comparison against another model				✓	✓
Generic Pattern: Showing pattern for each generated data stream matches observed patterns	✓	✓	✓	✓	✓
Characteristic Matching: Showing for each generated output data stream that it has correct seasonal or daily pattern		✓	✓	✓	✓
Relative Timing of Peaks: Showing time between peaks for dif. data streams matches observed dif.		✓	✓	✓	✓
Empirical Pattern: Showing pattern for each generated data stream matches empirical pattern – best for input streams			✓	✓	✓
Within Bounds: Showing for each generated output data stream that the mean of simulated stream falls within min/max of that stream for real data			✓	✓	✓
First moments: Showing for each generated output data stream that mean is not statistically different than real data – yearly, monthly or daily				✓	✓

Anthrax: BioWar vs. IPF based on Time to Death

Sverdlovsk

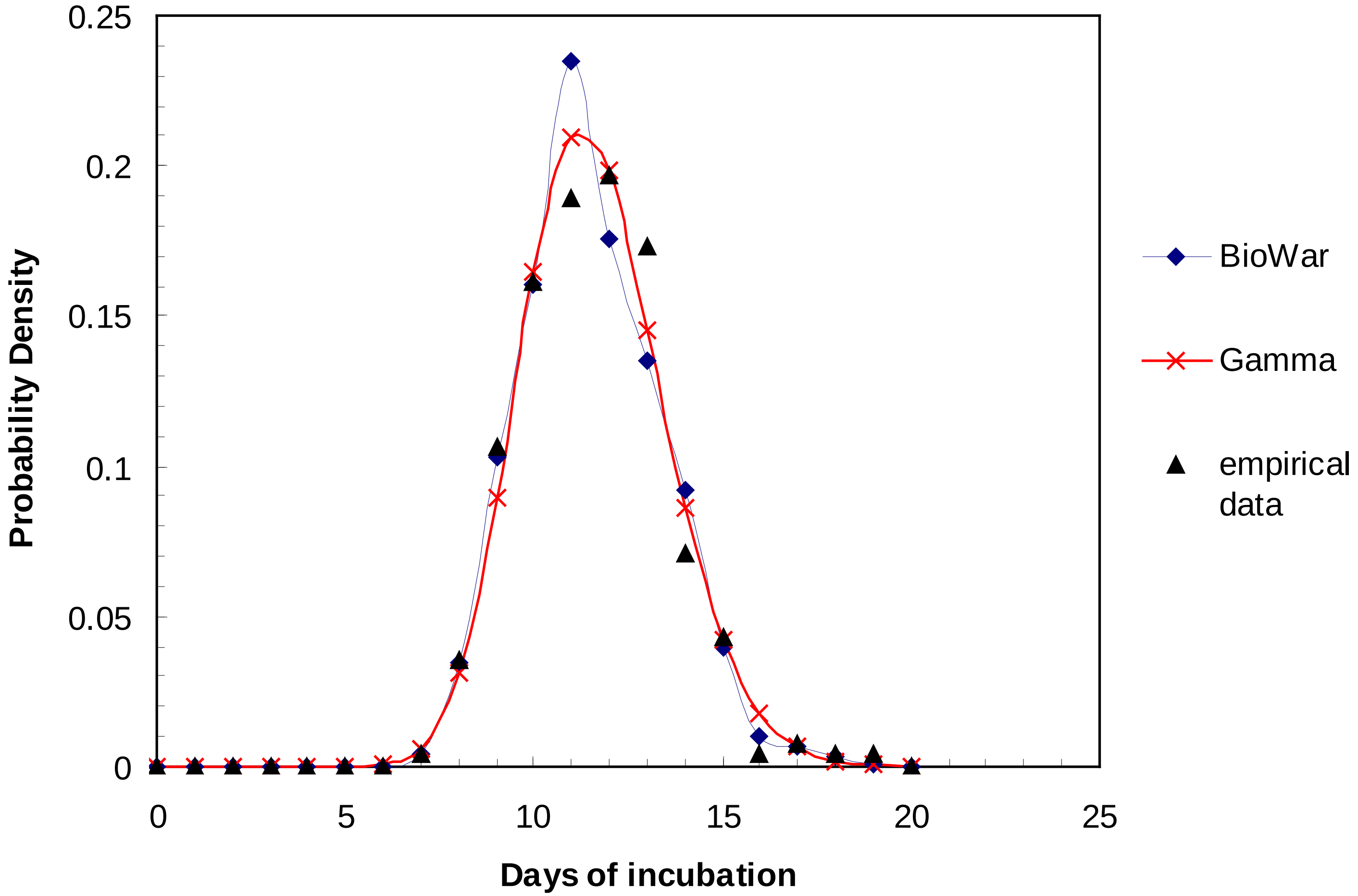


US Mail

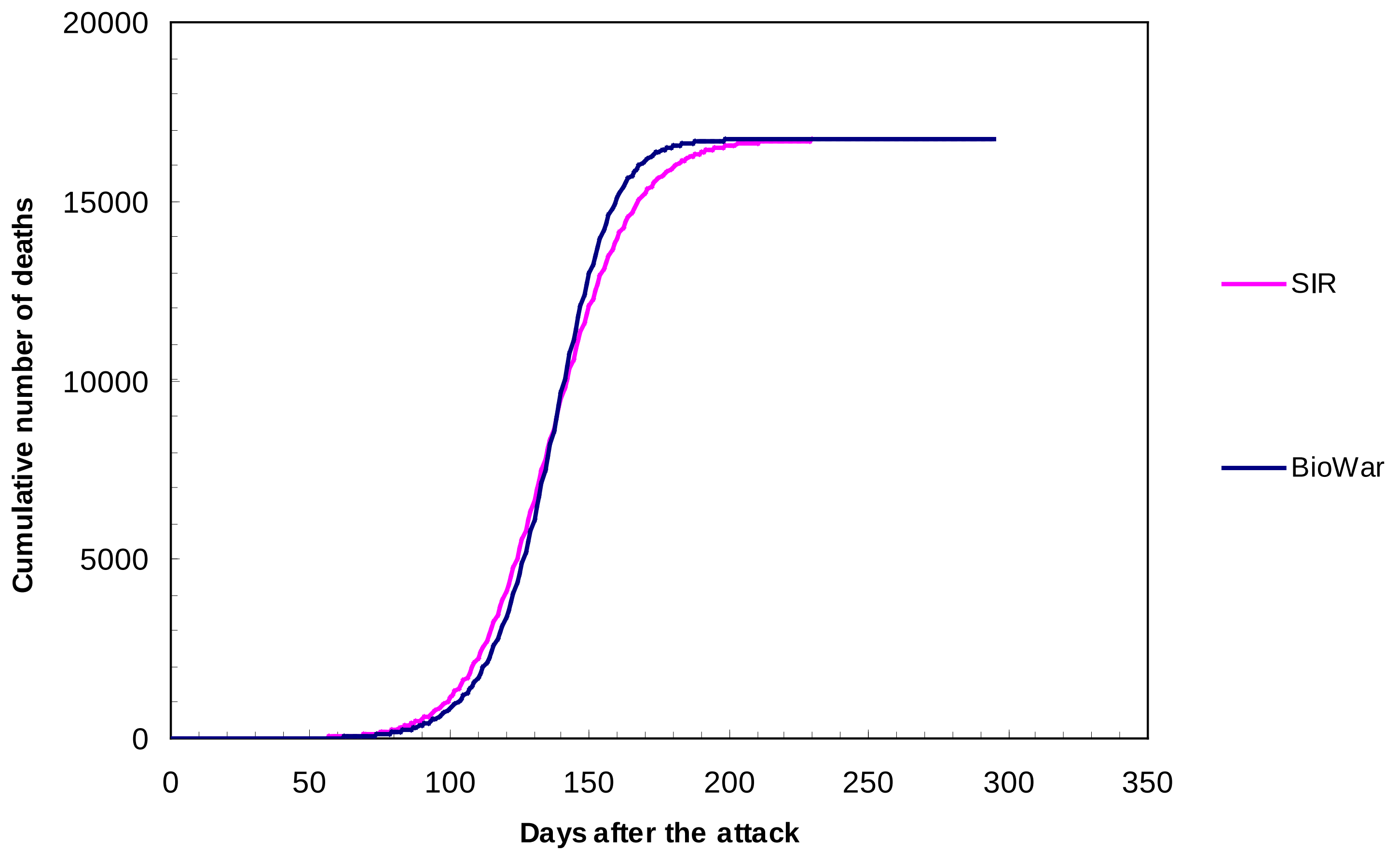


Carnegie Mellon

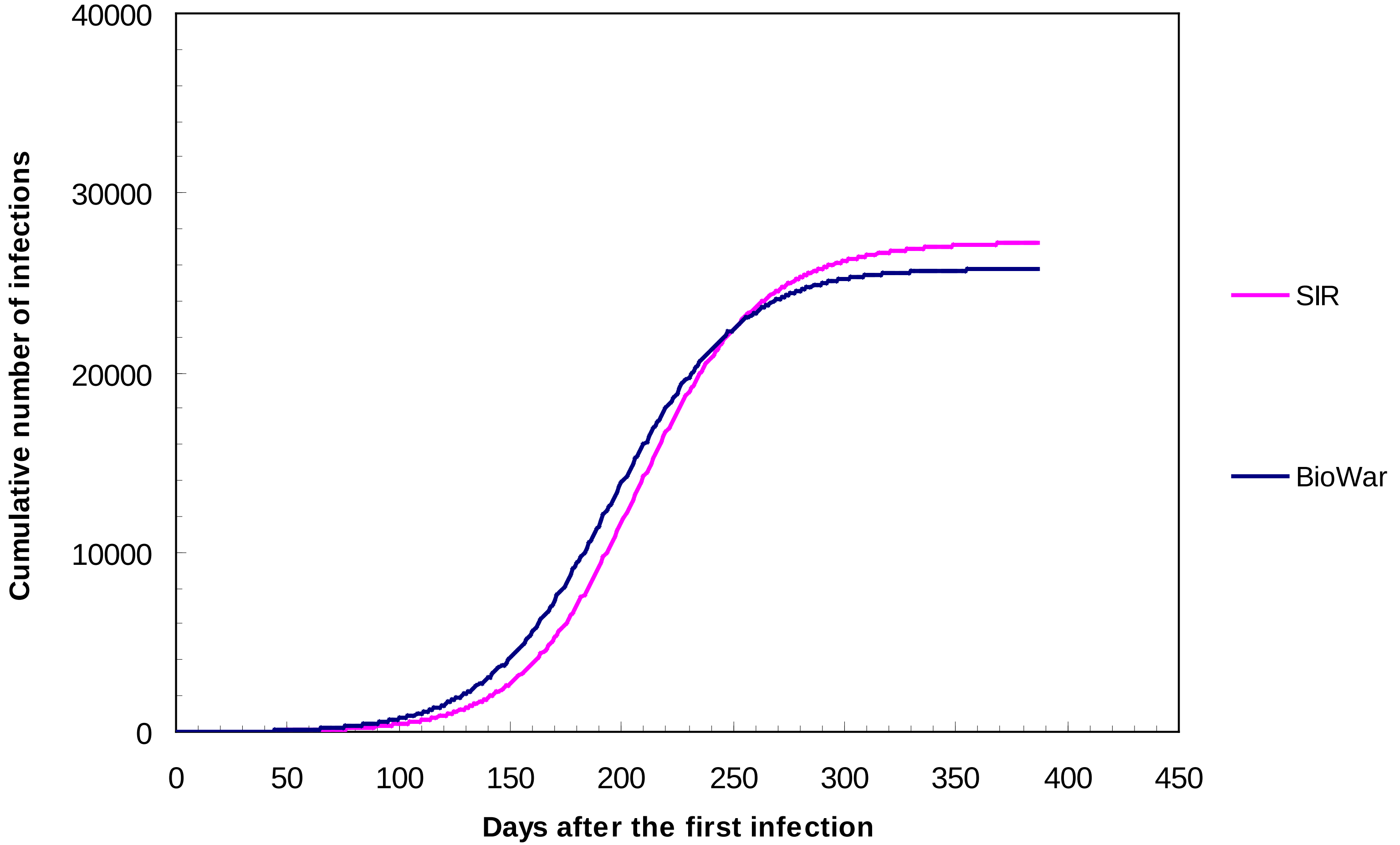
Smallpox Incubation



Smallpox Death: SIR vs. BioWar



Smallpox Infections: SIR vs. BioWar



Carnegie Mellon

Summary

- Social network multi-agent system
- Data for six metro-areas (census (business and population), weather)
- 562,000 agents, 60 diseases, 2 years, 10,000s locations, 6.5 hours, 4 processors
- Length of run is a function of the number of agents only
- Geometry grid – latitude/longitude, UTM
- Data reporting lags (based on real data)
- Existing output streams (CSV format)
    - Over the counter purchases (5 categories)
    - School/work capacity & attendance
    - Web lookups and phone calls
    - Doctor, ER (matches real for flu)
    - Epidemiological (EPI) curves per disease
- Validation at mean level for each output behavior

Current and Planned Enhancements

- Performance enhancements
- Simplified operation
- Training front end (Mass Casualty Model for Hampton MMRS) with:
  - Chemical attacks
  - Surge capacity
  - First responders
  - Dynamic parameter update

Performance Enhancements

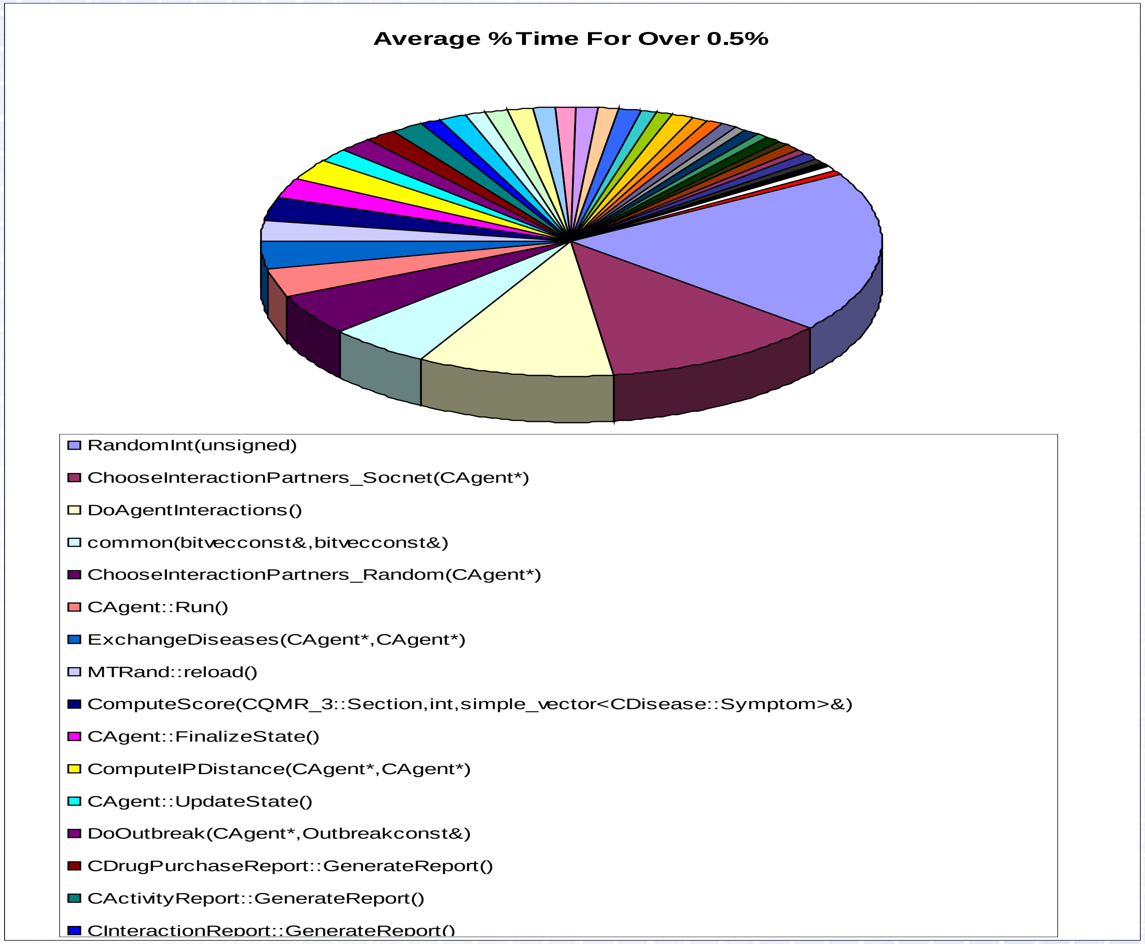
- Characterize existing performance
- Rework high overhead code
- Employ parallel execution
  - Multithreading
  - Multiprocessors
  - Grid computing

Current Performance Bottlenecks

- Profiling has indicated that BioWar spends the majority of its time in:
  - Computing Agent Interactions
  - Generating Random Numbers
  - Generating Reports

Carnegie Mellon

Current BioWar Performance Characteristics



CASOS

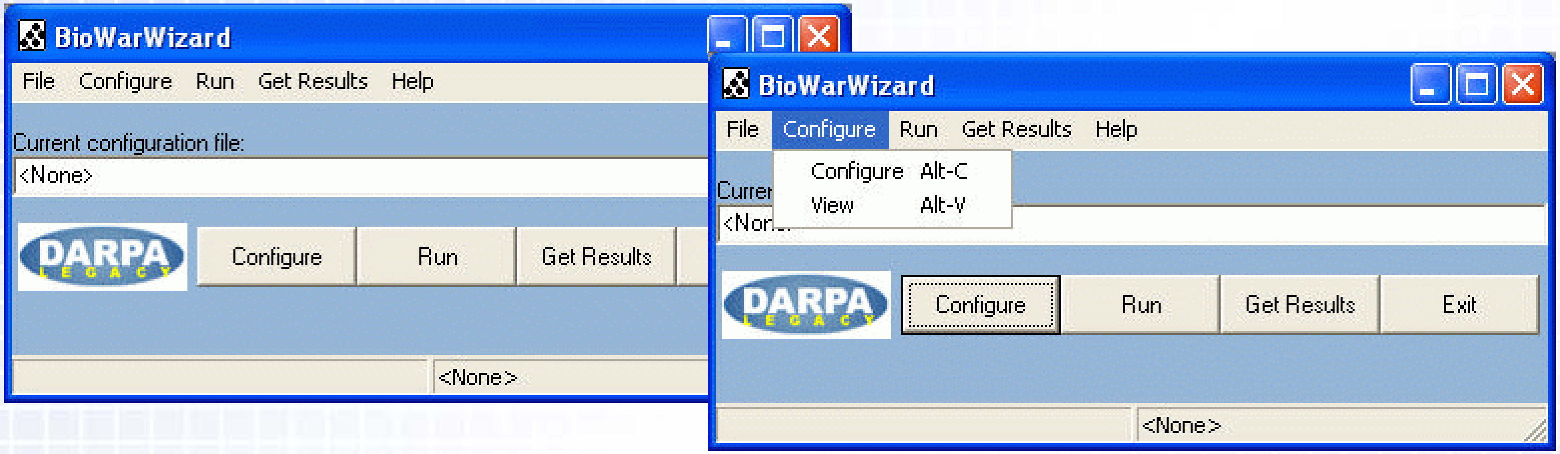
Plans for Performance Enhancement

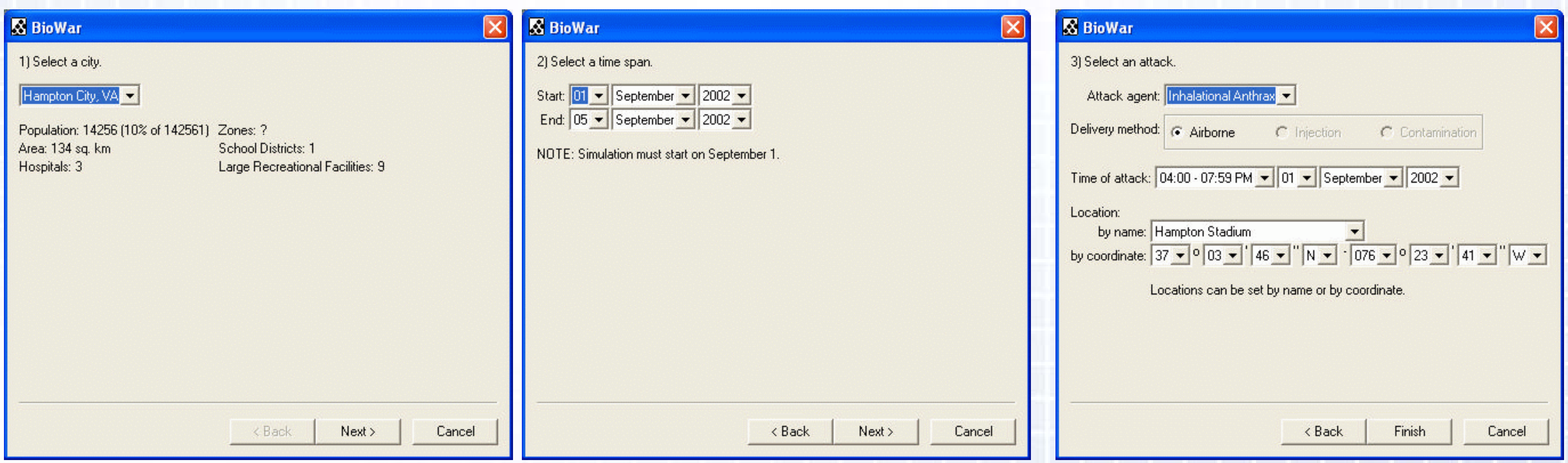
- Split the agent list between multiple threads
  - Most agent operations are autonomous and need little information about other agents to be executed
- Give each thread its own random number generator
- Distribute report generation between multiple threads
- Design threads that could be run as remote processes in the future

Simplified Operation -- Remote Interface to BioWar

- Proof of concept demonstration.
- Designed to allow user to:
  - Specify type and timing of a biological attack
  - Retrieve results.
- Client – server application
  - Interface runs on user’s system
  - BioWar runs on CMU host

Configuring BioWar





Checking the Configuration

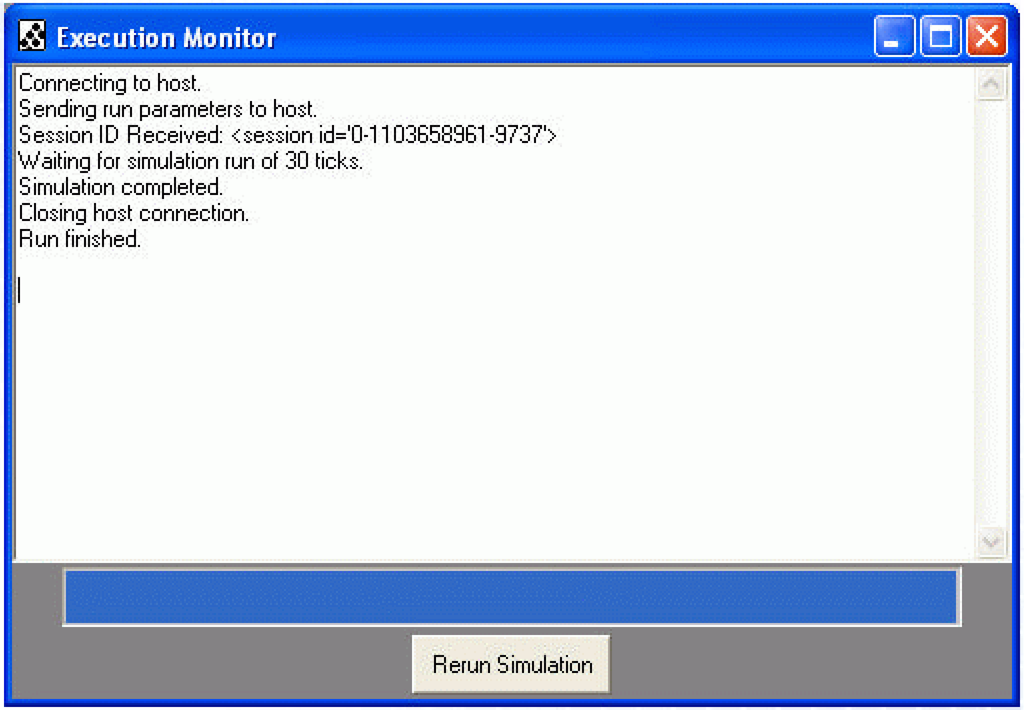
Configuration Viewer

Number of parameters = 24

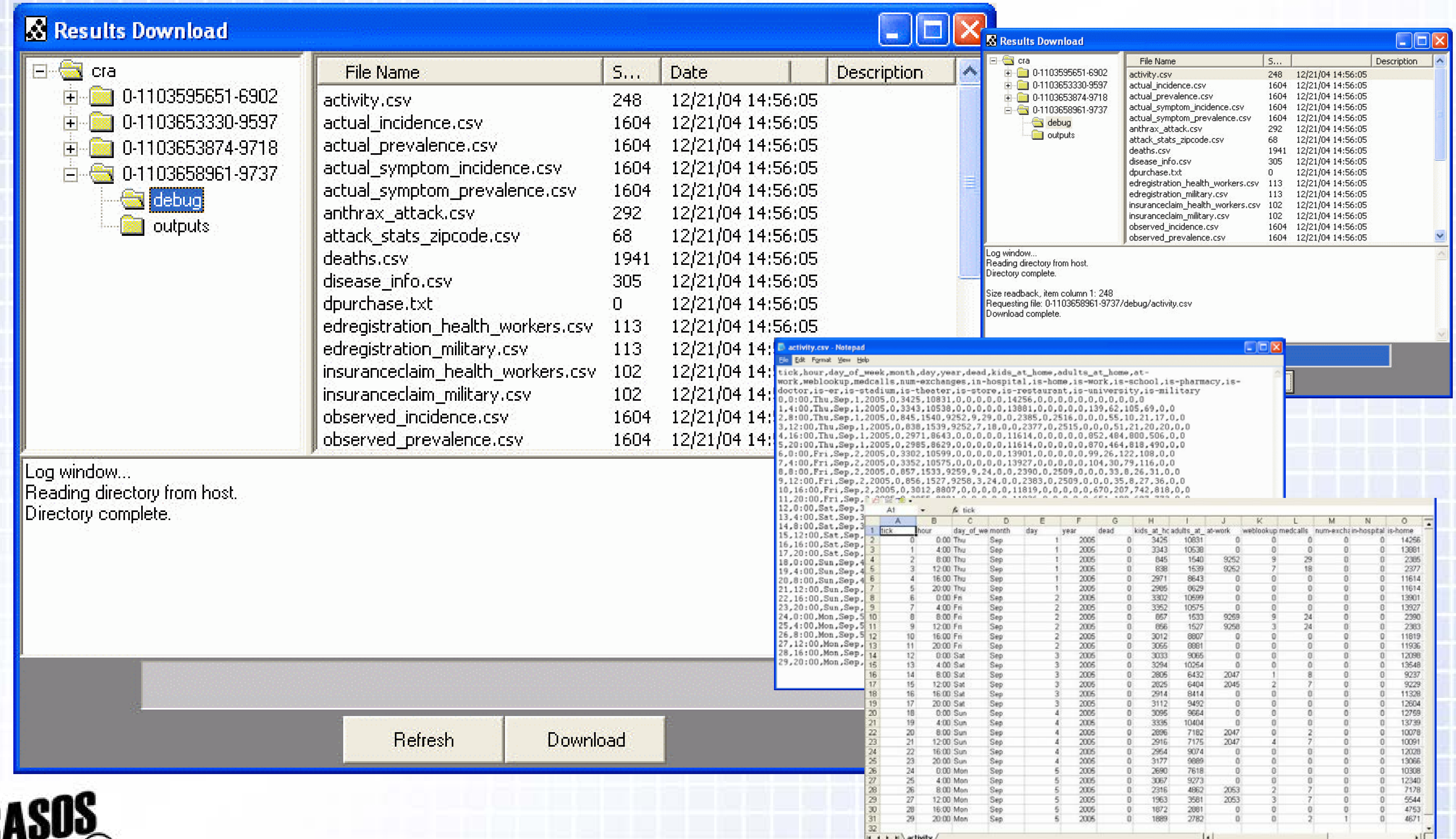
Type	Parameter Name	Current Value	Index	Display Text	Altered
0	CITY	hampton	0	Hampton City, VA	FALSE
1	SIM_START_DAY	1	0	01	FALSE
2	SIM_START_MONTH	9	0	September	FALSE
3	SIM_START_YEAR	2002	2	2002	FALSE
4	SIM_END_DAY	5	4	05	FALSE
5	SIM_END_MONTH	9	8	September	FALSE
6	SIM_END_YEAR	2002	2	2002	FALSE
7	ATTACK_DISEASE	anthrax_inhalational	0	Inhalational Anthrax	FALSE
8	DELIVERY_TYPE	airborne	0	Airborne	FALSE
9	ATTACK_TICK	4	4	04:00 - 07:59 PM	FALSE
10	ATTACK_DAY	1	0	01	FALSE
11	ATTACK_MONTH	9	8	September	FALSE
12	ATTACK_YEAR	2002	2	2002	FALSE
13	ATTACK_LOC_NAME	37.0629815619/-76.3948884096	1	Hampton Stadium	FALSE
14	ATTACK_LOC_LATITUDE	37.0629815619	0	<Attack Lat.>	TRUE
15	ATTACK_LOC_LONGITUDE	-76.3948884096	0	<Attack Long.>	TRUE

Refresh

Running BioWar



Retrieving Results



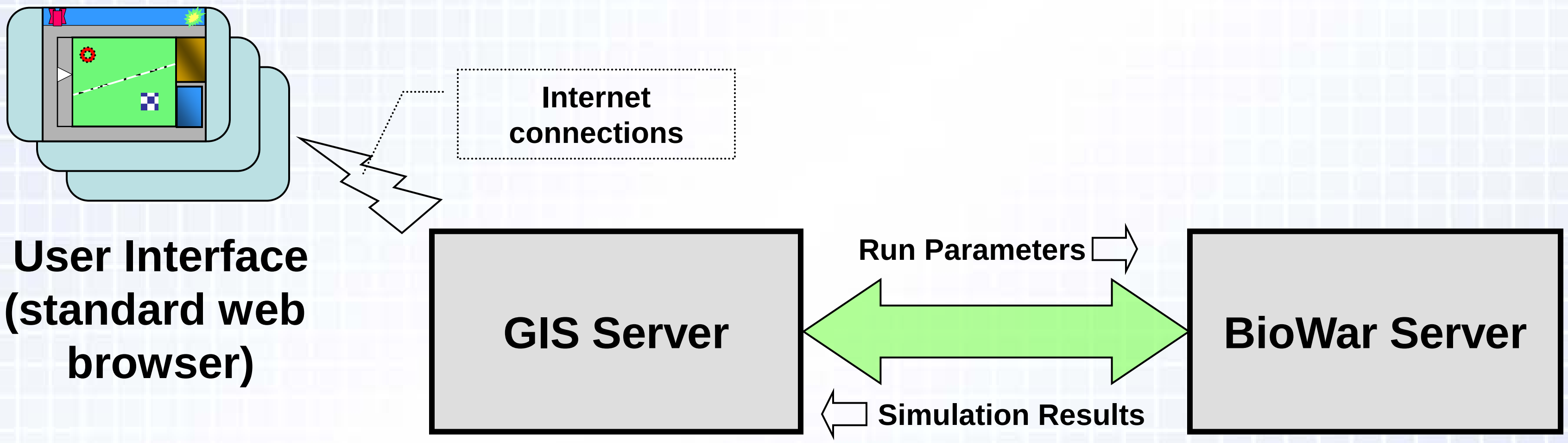
Hampton Roads Metropolitan Medical Response System

- Training aid for Metropolitan Medical Response System (MMRS) personnel
  - Graphical front end (HLS-GIS – Virginia Modeling, Analysis and Simulation Center, Old Dominion University)
  - Simulation back end (BioWar – CASOS, CMU)
- Will allow hospital, public health, and emergency managers to train for and to analyze mass casualty events.

MMRS Mission

- "Supports local jurisdictions' enhancing and maintaining all-hazards response capabilities to manage mass casualty incidents during early hours critical to life-saving and population protection,to include:
  - Terrorist acts using WMD/CBRNE
  - Large scale HazMat incidents
  - Epidemic disease outbreaks
  - Natural disasters"

System Structure (projected)



Additions to BioWar's Capabilities for MMRS

- Chemical attacks – add chemical effects into existing agent model
- Surge capacity – more realistically model hospital capacities in extraordinary circumstances
- First responders – specifically recognize their interaction with the injured and ill
- Dynamic parameter upgrade – allow human decisions to modify BioWar *during* a run

Chemical Attacks

- An agent based simulation simplifies addition of chemical weapons
- Chemical weapon effects differ from biological attacks
  - Smaller time gap between exposure and symptom expression
  - Potential for massive incidents
  - Relatively modest secondary “infection” effects
  - Response and contamination/decontamination protocols are different

Candidate Chemical Agents

Agent Type	Examples
Nerve Agents	Sarin (GB), Tabun (GA), Soman (GD), Cyclohesyl Sarin (GF), VX
Vesicants/Blister Agents	Sulfur Mustard, Lewisite, Nitrogen Mustard, Mustard Lewisite, Phosgene-oxime
Pulmonary/Choking Agents	Phosgene, Chlorine, Diphosgene, Chloropicrin, Oxides of Nitrogen, Sulfur Dioxide
Cyanides	Hydrogen Cyanide (HCN), Cyanogen Chloride
Lacrimating Agents	Pepper gas, Chlorpacetophenone
Vomiting Agents	Adamsite

Initial BioWar Chemical Capacity

- Selected chemical agents for implementation in BioWar:
  - Sarin – war agent
    - Actual events for comparison (Aum Shinrikyo, Japan)
    - Fairly well studied
  - Chlorine – industrial chemical
    - Used in large quantities commercially
    - Release may occur deliberately or accidentally
- Three areas affected in BioWar:
  - Attack and exposure modeling (delivery, puff/plume modeling, dose, and absorption modes)
  - Effect on individual agents (symptoms, progression and lethality)
  - Medical response (diagnosis, treatment protocol, recovery and surge capacity)

Current Status of Implementation

- Characterized Sarin
  - Lethality
  - Symptoms
  - Treatment
- Defined wind models
  - Puff – explosive/mass release incidents
  - Plume – storage leaks/industrial accident scenarios
- Prototyped key routines
  - Puff release
  - Sarin exposure
  - Sarin leathality
  - Sarin symptoms
- Researching
  - Long term effects of nerve agents
  - Chlorine
- Revising BioWar structure
  - Chemical attack insertion
  - Configurable response
  - Key medical supplies and capacity limits

Carnegie Mellon

Selected Sarin Properties

Military Classification	Lethal Agent (Nerve Gas)
Most Likely Form to Be Disseminated	Vapor, Aerosol, or Spray
Types of Weapon Suitable for Disseminating the Agent	All Types of Chemical Weapon
Maximum Weight Delivered by a Light Bomber (4-ton bomb load)	1000 kg
Approximate Solubility in Water at 20ºC	100%
Volatility at 20ºC	12 100 mg/m³
Physical State (a) at -10ºC (b) at 20ºC	Liquid Liquid
Duration of Hazard (a) 10ºC, rainy, moderate wind (b) 15ºC, sunny, light breeze (c) -10ºC, sunny, no wind, settled snow	(a) ¼ h – 1 h (b) ¼ h – 4 h (c) 1 – 2 days
Casualty-producing Dosages	> 5 mg-min/m³
Estimated Human Respiratory LC(50)	100 mg-min/m³
Estimated Human Lethal Percutaneous Dosages	1500 mg-min/m³

CASOS

Nerve Agents: Syndromes

The three levels of symptoms according to the dosage received:

I. Low Exposure:
- Pinpoint pupils (miosis)
- Bronchoconstriction
- Respiratory arrest
- Hypersalivation
- Increased secretions

II. Medium Exposure:
- The above plus
- Diffuse muscle cramping
- Muscle tremors
- Dimming of vision and eye pain
- Generalized weakness
- Severe headache
- Confusion
- Drowsiness

III. High Exposure:
- The above plus
- Involuntary urination and defecation
- Very copious secretions
- Twitching, jerking, staggering, and convulsions
- Sudden loss of consciousness
- Seizures
- Flaccid paralysis
- Coma

Carnegie Mellon

Gaussian Puff Dispersion Model

$$C(x,y,z,t) = \frac{Q}{\sqrt{2}\pi^{3/2}\sigma_x\sigma_y\sigma_z}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-ut}{\sigma_x}\right)^2 + \frac{y^2}{\sigma_y^2} + \frac{(z-H)^2}{\sigma_z^2} + \frac{(z+H)^2}{\sigma_z^2}\right]\right\}$$

where

C - concentration of pollutant;

Q – fixed mass of poisoning material;

u – air velocity;

H – height of puff release;

t – time;

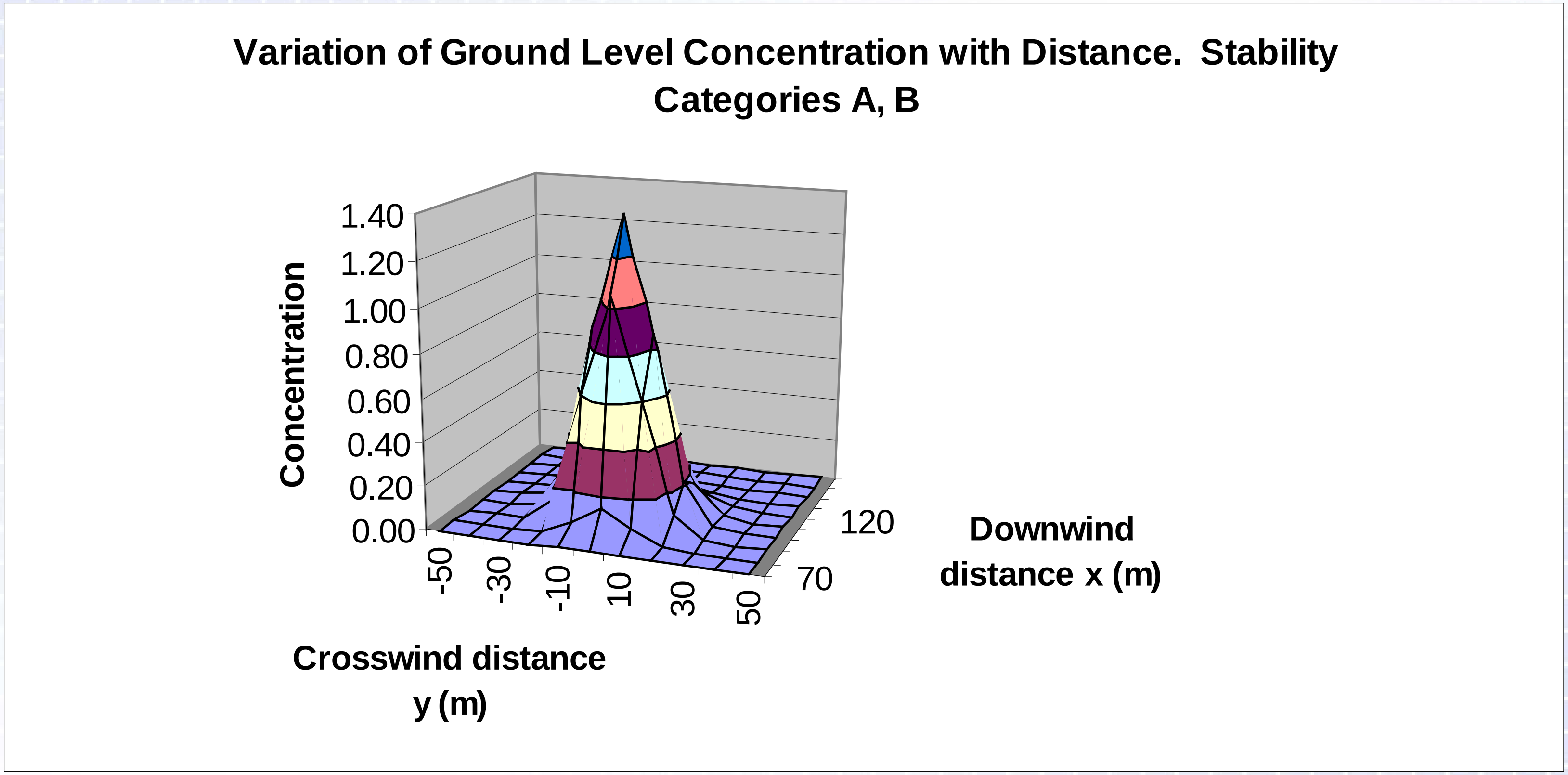
x, y, z – distance from the release;

$\sigma_x, \sigma_y, \sigma_z$ - standard deviations along the axis

Atmospheric Stability Categories

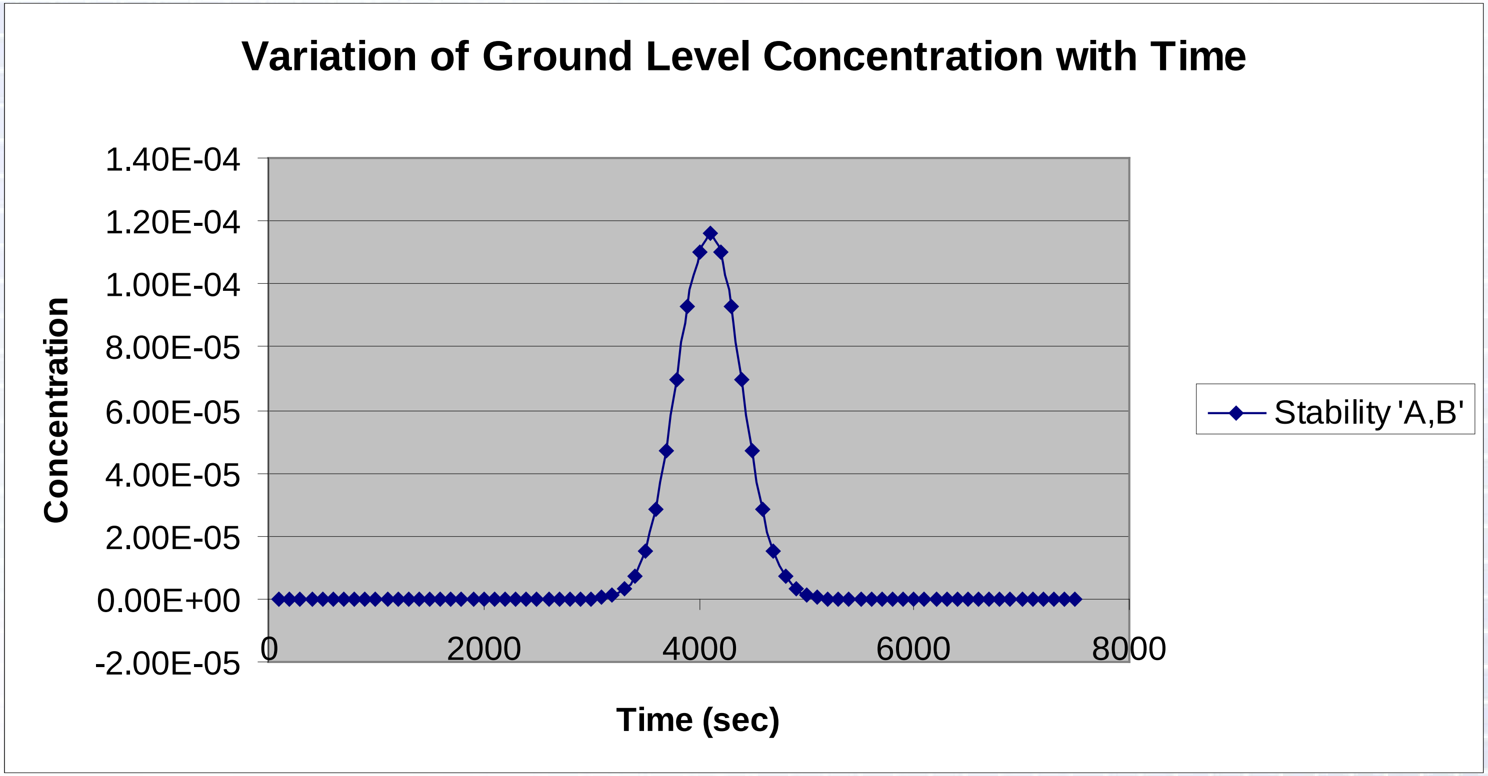
Stability Category	Classification	Natural Phenomena	Most Likely Occurrence
A	Extremely Unstable	**Strong thermal instability, bright sun**	**Late morning to mid afternoon in spring and summer**
B	Moderately Unstable	**Transitional periods, moderate mixing**	**Day time transitions all year**
C	Slightly Unstable	**Transitional periods, slight mixing**	**Day time transitions all year**
D	Neutral	**Strong winds, overcast, day/night transitions**	**Day time/cloudy; night time/cloudy; high wind**
E	Slightly Stable	**Transitional periods, night time moderate winds**	**Night time transitions all year**
F	Moderately Stable	**Clear night time skies, very limited vertical mixing**	**Night time, clear skies, light winds all year**
G	Extremely Stable	**Plume fanning and meandering**	**Night time, clear skies, no wind all year**

Sarin Modeling Results - I

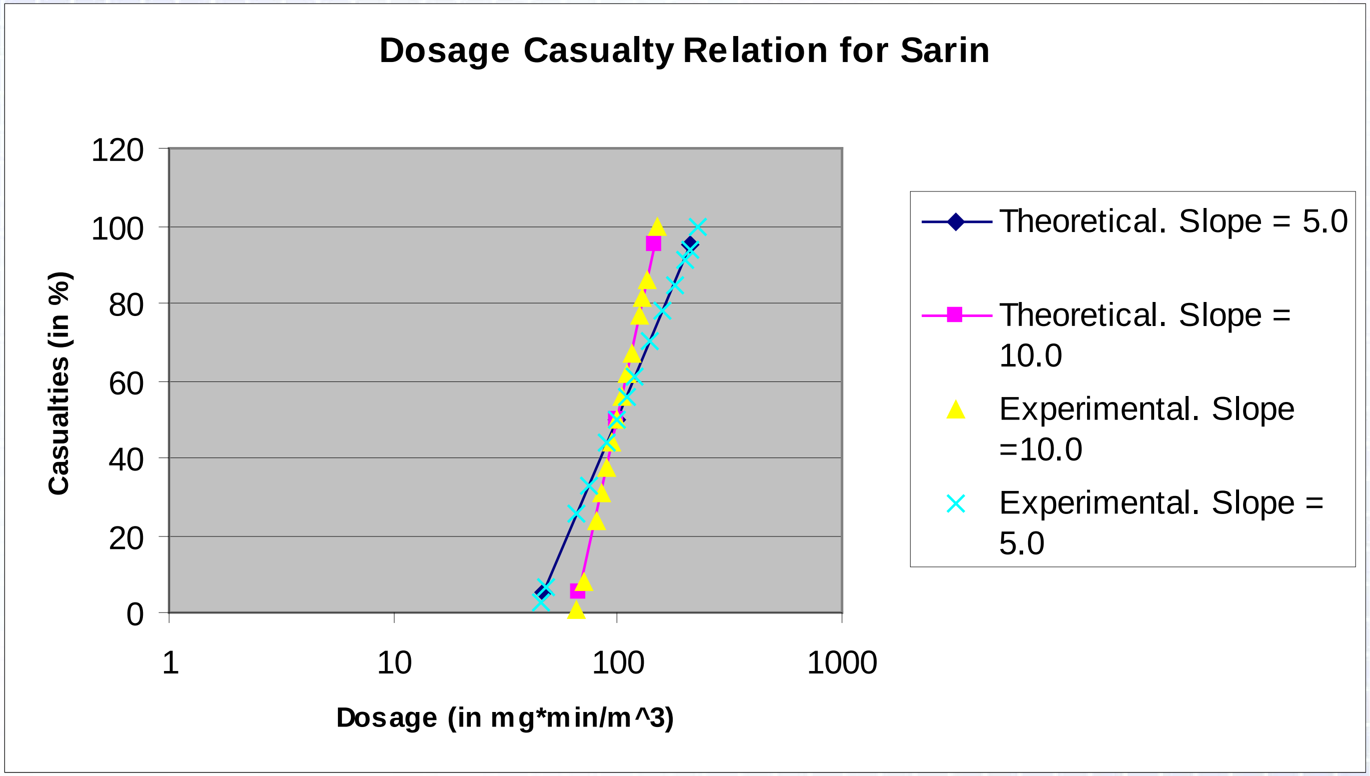


Carnegie Mellon

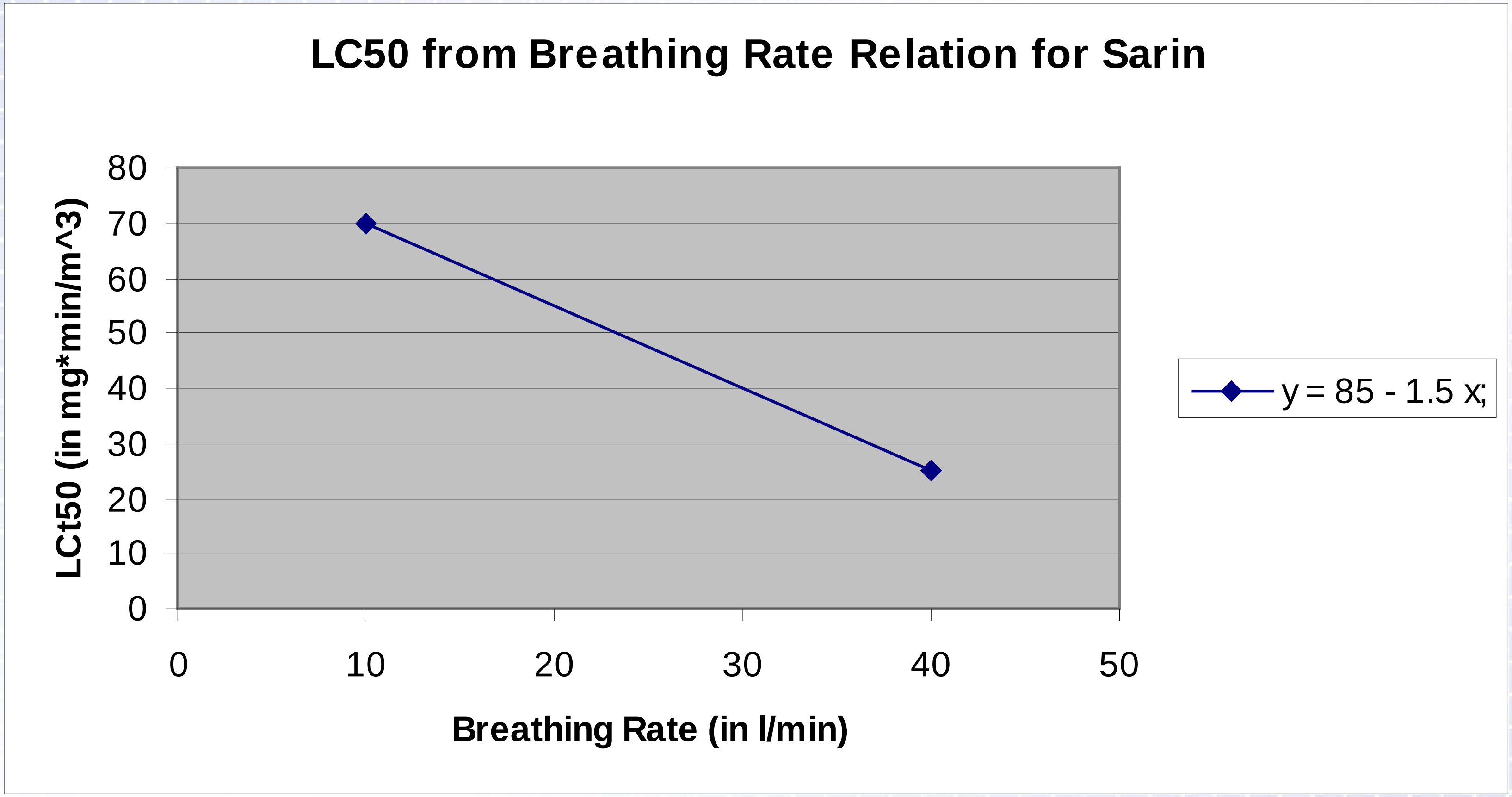
Sarin Modeling Results - II



Sarin Modeling Results - III



Sarin Modeling Results - IV



Nerve Agents: Treatment

Hospital Treatment for Inhalation and Dermal Absorption

- Atropine (2 mg) IV; repeat q 5 minutes, titrate until effective, average dose 6 to > 15 mg. Use IM in the field before IV access (establish airway for oxygenation)
- Pralidoxime chloride (2-PAMCl) 600 – 1800 mg IM or 1.0 g IV over 20-30 minutes (maximum 2 g IM or IV per hour)
- Additional doses of atropine and 2-PAMCl depending on severity
- Diazepam or lorazepam to prevent seizures if > 4 mg atropine given
- Ventilation support

Field Treatment:

- Mark I kit: was designed for military self-administration in the field. It consists of 2 spring-loaded devices to inject yourself, containing atropine and pralidoxime. Are not yet available for civilian use.

Selected Chlorine Properties

- Chlorine is a greenish yellow gas
- It has a strong, pungent odor; because its odor threshold of 0.08 ppm is below the level of toxicity, it is adequate warning
- It is heavier than air in its pure form
- It is an oxidizing agent that is highly reactive with water and liberates hypochlorous acid, hydrochloric acid, and oxygen-free radicals, which are toxic to tissues
- It has intermediate solubility
- It is also soluble in alkalis, alcohols, and chlorides
- It is not combustible, but as an oxidizer, it may react violently with many materials including fuels

Chlorine Inhalation Toxicity

Inhalation toxicity is a function of the dose received and is dependent on the concentration of gas and duration of exposure

- The permissible exposure level is 1 ppm
- The minimum concentration immediately dangerous to life and health is 25 ppm
- Exposure to more 50 ppm is dangerous
- Exposure to 1000 ppm is fatal even with short exposures
- The lowest reported lethal concentration is 430 ppm for 30 minutes

Chlorine Syndromes

Signs:

- Pulmonary edema with some mucosal irritation leading to acute respiratory distress syndrome (ards)
- Pulmonary infiltrate
- Violent cough
- Nausea and vomiting
- Lightheadedness and headache
- Chest pain or retrosternal burning
- Muscle weakness
- Abdominal discomfort

Symptoms:

- Shortness of breath
- Chest tightness
- Wheezing
- Laryngeal spasm
- Mucosal and dermal irritation and redness

Gaussian Plume Dispersion Model

$$C(x,y,z,t)=\frac{Q}{2\pi u\sigma_y\sigma_z}\exp\left[-\frac{y^2}{2\sigma_y^2}\right]\left[\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right)+\exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right]$$

where

C - concentration of pollutant;

Q – fixed mass of poisoning material;

u – air velocity;

H – height of puff release;

x, y, z – distance from the release;

$\sigma_x,\sigma_y,\sigma_z$ **- standard deviations along the axis**

Carnegie Mellon

Pulmonary/Choking Agents: Onset and Treatment

Onset

- 1 - 24 hours (rarely up to 72 hours);
- May be asymptomatic period of hours

Treatment of Inhalation Cases

- No antidote
- Management of secretions
- Oxygen therapy
- High dose steroids to prevent pulmonary edema
- Ventilatory support
- Treat pulmonary edema with positive end-expiratory pressure (PEEP)

Surge Capacity

- BioWar originally focused on support event detection
  - Accurate steady state model
  - Generation of realistic reports
- Enhance to simulate emergency needs
  - Limit hospital capacity appropriately
  - Augment medical capacity in emergencies
    - Increase existing locations
    - Activate field hospitals and clinics
  - Supply depletion and replenishment
    - Model critical emergency supplies flexibly
    - Decrement as a consequence of treatment
    - Allow emergency shipments, mass treatment campaigns

Surge Capacity Implementation

- Capacity
  - Using MMRS guidelines for capacity
  - Hospital capacity corresponds to staffing
  - Emergency locations preplanned according to emergency plans
- Supplies
  - Focus on emergency supplies
  - Both expended (drugs) and reusable (ventilators) considered

First Responders

- First responders and medical staff should have increased contact with the sick and injured (with the chance of infectious disease transmission).
- BioWar uses agent to agent interaction to model infectious disease transmission.
  - Medical staff who work in fixed medical locations will <u>automatically</u> interact with the ill in BioWar.
  - First responders do not (always) work from a fixed medial locations and will not automatically interact.
- Solution – add a special interaction category to simulate first responder contact.

Interaction Algorithm

An interaction list is built every tick for each agent based on:

- **Social network (modified by time of day)**
- **Random selection (within a specified distance, so co-location is significant)**

Interactions are resolved, including disease transfer.

First responder enhancement:

1. At simulation start, assign agents as first responders.
2. Every tick, make a list of available first responders.
3. Ill/injured agents heading to the emergency room may require first responder assistance.
4. Add one or more first responders to the ill agent's interaction list.
5. Resolve interactions normally.

Who Is a First Responder?

- First responders are defined using the Standard Occupational Classification (SOC) job categories
  - 29-2041 Emergency Medical Technicians and Paramedics
  - 33-1012 First-Line Supervisors/Managers of Police and Detectives
  - 33-1021 First-Line Supervisors/Managers of Fire Fighting and Prevention Workers
  - 33-2011 Fire Fighters
  - 33-2021 Fire Inspectors and Investigators
  - 33-3051 Police and Sheriff's Patrol Officers
  - 33-3052 Transit and Railroad Police
  - 53-3011 Ambulance Drivers and Attendants, Except Emergency Medical Technicians
- Employment levels from the Occupational Employment Statistics (OSE) survey
- *Would like to add volunteer first responders*
  - *No national database found*
  - *BioWar does not have secondary/volunteer jobs (agents map 1:1 to jobs)*
  - *Information would ideally be specified by county.*

When are First Responders Needed?

- Some agents going to a hospital emergency room need first responder assistance.
  - Agent must have “customer intent”.
  - Agent was not in the ER last tick.
  - First responder request rate is 25%
- What is actual rate of first responder use for ER visits (25% is arbitrary)?
- What other first responder interactions are important?

Dynamic Parameter Upgrade

- BioWar already enables batch execution
- Need to add support waiting for human decision
  - Pause to wait for human participants
  - Dynamic configuration update, implementing decisions
- Pause capacity and dynamic parameter input has been demonstrated using current BioWar versions

Dynamic Parameter Upgrade Purpose

- Support adjustment of critical parameters types:
  - Responds to simulation events – attacks, timed events, disease detection…
  - Allows adjustment of critical parameters – vaccination, supply levels, preparedness, agent actions…
  - Allows selection of specific populations and locations – all doctors, hospital #3, agents within 1,000 meters of an attack…
- Enable both batch and dynamic operation

Response Syntax

- Defining a way to specify new parameters (responses)
- General form of responses:

  <trigger> <target> <response>

```
on(<trigger>) [wait] for(<target>) do(<response>)
```

- Example: On May 13, decrease the supply of Cipro by 35%

```
on(5/13) for(supply types(<CIPRO>)) do(*0.65)
```


Key Modules

1. City Description Construction
2. Social Network Construction
3. Medical (including Disease & Diagnosis)
4. Agent Behavior Module
    1. Interaction
    2. General Action – recreate, work, school
    3. Self Diagnosis – OTC, Dr., E.R.
    4. Etc.
5. Geometry (including geometric grid)
6. Weather (including Wind & Climate)
7. Dispersion for Aerosolized Attacks
8. Attack Scenario Generation
9. Post Processors
10. Automated Validation: WIZER

Module 1: City Description Construction

- City construction initializes geometry, demographics, agent definitions, and other inputs for one or more simulation runs
- Process overview (“gensim”):
  - Load configuration file (sets city scale, calendar range, etc.)
  - Load global data (disease data, generic statistics, etc.)
  - Load city data (geometry, population & school demographics, location positions and sizes, weather system, attack and outbreak specs, etc.)
  - Generate city (random population, agent social network, locations, jobs & schools, outbreak and attack calendar, weather calendar, etc.)
- Inputs generated for simulator (“BioWar”)
  - Population and infrastructure (properly-distributed agents, jobs, schools, entertainment locations)
  - Simulator data (social network, weather calendar, attack & outbreak instance characteristics)
- Note: generated cities are saved enabling multiple runs on same city

Carnegie Mellon

Module 1: City Construction (Cont'd)—Random Population

- Per-census tract demographics for simulation region are extracted from locally installed Census database
- Agents are assigned to tracts by throwing a random die *R* in [0,1] against a cumulative probability distribution over tract population
- Agent profiles are similarly generated using cumulative distributions over the per-tract demographic profiles

Module 1: City Construction (Cont'd)—Ego Net and Home Generation

- Census tract assignments are then used to guide ego net generation for families
- Family ego nets are used to determine "cohabitating agents"
- Cohabitating agents are then assigned to the same home location (affects job and school assignment)

Module 1: City Construction (Cont'd)—School Generation and Assignment

- Per-school demographics are extracted from a locally-installed NCES CCD dataset (public schools only)
- Schools are mapped to the districts that contain them (schools are guaranteed to have a district)
- Agents with homes in each school district are then assigned randomly by age to a school in that district

Module 1: City Construction (Cont'd)—City Infrastructure Generation

- City infrastructure currently consists of locations only (some random, some specified by a GNIS database)
- Positions of jobs, doctors, pharmacies, restaurants, stores, & theaters generated randomly, with capacities from a NAICS database
- Positions of hospitals and parks from GNIS, capacities from NAICS
- Randomly generated positions are distributed uniformly over the city based upon census tract population and geometry
- Agents are currently randomly assigned to jobs

Carnegie Mellon

Module 1: City Construction (Cont'd)—Other Generated Inputs

- Weather calendar (described later) includes wind (for attack resolution) and climate (temperature, pressure, precipitation)
- Schedule for outbreaks and attacks based upon user-specified parameters (also described later)

CASOS

Module 1: Data Sources for City Construction

- Urban Area Definitions
  - US Census Bureau - Metropolitan Statistical Areas: 1999
  - ZIP Codes
    - ZipExpress™ – Lookup Zip Codes by County
    - Capitolimpact.com – Capitolimpact Gateway
- US Census Bureau Cartographic Boundary Files
  - MSA boundaries
    - Census Tracts: 2000
    - 3-Digit ZIP Code Tabulation Areas (ZCTAs): 2000
    - 5-Digit ZIP Code Tabulation Areas (ZCTAs): 2000
  - Schools
    - School Districts - Elementary: 2000
    - School Districts - Secondary: 2000
    - School Districts - Unified: 2000
- Location Names, Counts and Geographic Coordinates
  - US Census Bureau - 2000 Economic Census
  - NAICS – count of entertainment/recreation, work, doctor, pharmacy locations
  - GSS – ego net input (indirectly affects number of homes)
  - USGS – GNIS (Geographic Names Information System) – positions of hospitals, parks & stadiums
  - NCES – CCD Public School District Data
  - NCES – CCD Public School Data

Module 1: City Locations

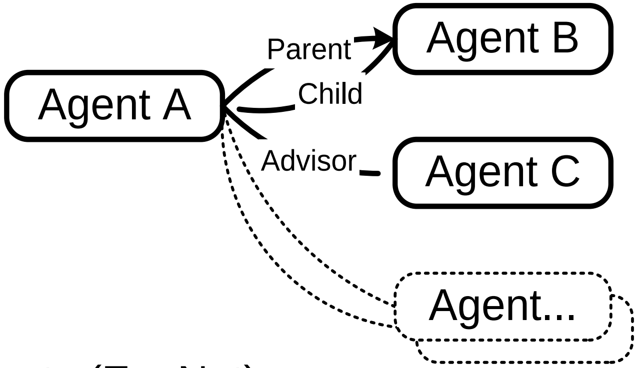
	Pittsburgh	San Diego	Norfolk	Hampton*
Doctors	1951	1776	841	75
Hospitals/ER	50	33	19	3
Pharmacies	479	274	199	16
Restaurants	4383	4886	2504	203
Stadiums	200	143	97	10
Stores	7540	8109	4944	374
Theaters	551	516	307	30
Population	2,358,695	2,813,833	1,569,541	146,431

Module 2: Social Network Construction

- Predefine for each agent who is in their ego net
  - Set of others they primarily interact with
- Set this based on empirical data on size and constitution of networks
- This is a limit on the set of others an agent interacts with on average
- Key to having a realistic synthetic population
- Note: generated networks are save enabling multiple runs on same population

Module 2: Social Network Construction

- Social network describes established human relationships
  - Family
    (spouse, parent, child, sibling, other family)
  - Proximity based
    (coworker, schoolmate, group member, neighbor)
  - Voluntary
    (friend, advisor, other)



- Each agent has list of connections to other agents (EgoNet):

  [(<agent> <relationship>)   [(<agent> <relationship>) …]]

- Factors considered during creation
  - Target network size for agent
  - Frequency of relationship type
  - Agent demographics
  - Agent’s customary locations (home, school, work)

Module 2: Validation and Tuning

Agent Social Network Size				
	Expected (From Klovdahl Study)	**Norfolk**	**San Diego**	**Pittsburgh**
Average Social Net Size	**33**	**28**	**28**	**28**
Range	**6-97**	**8-67**	**6-68**	**7-79**

Module 3: Medical Disease and Diagnosis

- Symptom based general model of disease
- Agents self diagnose on the bases of visible symptoms
- Prevalence of diseases based on CA data
- Medical personnel diagnose on the basis of visible and non-visible symptoms
  - Tests are employed
  - Tests vary in diagnostic accuracy
  - Tests vary in time to get report
  - Type 1 and 2 errors possible
- EPI Curves are an OUTPUT not an INPUT
  - Can be generated for observed and actual cases
  - Note: testing detection routines with diagnostics off may be misleading

Module 3: General Model of Disease

- Based on *symptoms* rather than unseen parameters like viral load
  - People change their behavior based on symptoms, not viral/disease parameters
  - With database of symptoms (and associated behavior changes), easy to construct arbitrary and new diseases (SARS)
- Parsimonious representation and calculation of symptom progression
- Can represent both contagious and non-contagious diseases
- Diseases have stochastic nature (not everyone presents like the textbook), and our model can represent outliers

Module 3: Disease Features

- Strain severity
- Onset locations
- A set of symptoms
  - Evoking strength, P(D|S) (where D=disease, S=symptom)
  - Frequency, P(S|D)
  - Cost of treatment (low, medium, high)
  - Visible or requiring test (tests are visible, low cost, high cost)
- Progression of disease within agent
  - Infected phase: agent has been infected but does not infect others
  - Communicable phase: agent infects others (only exists for contagious diseases)
  - Symptomatic phase: agent displays symptoms
    - Onset of specific symptoms is random
  - Variations in onset and length of each phase in general
    - Known timing (CDC or JH web sites)
    - Additional variation per agent based on
      - Severity of strain
      - Agent age, gender, race, medical history
      - Treatment

Module 3: Model for Contagious Disease

- Transmission medium: contact, airborne, food, etc.
- When person comes into the contact with the transmission medium, disease transmission occurs with some probability.
- Phases: infected, communicable, symptomatic
- Modeled, at least partially, as non-deterministic automata.
  - As past medical history affects the transition, this is a non-Markovian model.
  - Any time within the duration of a state an intervention can occur, and the reality changes.
  - The state of the disease can also affect the intervention, e.g, certain symptoms trigger certain behaviors.
- In contrast to SIR model, we model contagious diseases at the individual level and take intervention into account.

Carnegie Mellon

Module 3: List of Contagious Diseases

Bacterial pharyngitis acute non streptococcal non gonoccocal,
botulism,
bubonic plague,
campylobacter enteritis,
cutaneous atypical mycobacterial infection,
encephalitis acute viral,
giardiasis intestinal,
gram negative pneumonia non klebsiella,
hepatitis A acute,
herpes simplex encephalitis,
immunice deficiency syndrome acquired (aids),
infectious mononucleosis,
influenza,
influenza pneumonia,
malaria,
meningococcal meningitis,
mycoplasma pneumonia,
plague meningitis,
plague pneumonia,
pneumococcla pneumonia,
pulmonary legionellosis,
salmonella enterocolitis non typhi,
schistosomiasis systemic, shigellosis,
staphylococcal pneumonia,
staphylococcal scarlet fever
toxic shock syndrome,
streptococcal pharyngitis acute,
streptococcus pyogenes pneumonia,
syphilis primary,
smallpox,
tuberculosis chronic pulmonary,
tuberculosis disseminated,
varicella pneumonia,
viral gastroenteritis,
viral pharyngitis acute non herpetic

CASOS

Module 3: Model for Non-Contagious Disease

- Non-contagious disease do not have a communicable phase
- Some non-communicable can be spread by contact
  - E.g., anthrax spread by US Mail
  - Implementation of this is planned
- Modeled, at least partially, as non-deterministic automata.
- Intervention affects how states in the model change.
  - If anthrax infection is suspected to be present, this triggers the intervention such as giving Cipro antibiotics. Giving Cipro, in turn, ameliorates the possible symptoms and possibly cure the disease.
- For short term non-contagious
  - E.g., food poisoning
  - Outbreaks are randomly determined based on prevalence data
- For long term non-contagious diseases
  - E.g., angina, diabetes
  - Given prevalence information initial population is “infected”
    - Subject to known race, gender, age distributions
  - If an agent dies another agent at random is infected

Module 3: List of Non-Contagious Diseases

Angina pectoris,
anxiety neurosis,
arteriolar nephrosclerosis benign essential hypertension,
arteriosclerotic heart disease,
bronchial asthma,
bronchitis chronic simple,
brucellosis,
cardiogenic shock acute,
chronic fatigue syndrome,
cutaneous anthrax,
depression,
diabetes mellitus,
disseminated intravascular coagulation,
fibromyalgia syndrome,
heat exhaustion,
hypertensive heart disease,
hypovolemic shock,
anthrax inhalational,
myocardial infarction acute,
obsessive compulsive neurosis,
pulmonary emphysema,
somatization disorder hysteria,
staphylococcal gastroenteritis food poisoning, tension headache,
tularemia,
tularemia menigitis

Module 3: Weaponized Diseases

- Cutaneous Anthrax
- Inhalation Anthrax
- Smallpox
- Bubonic Plague

Module 3: Medical Diagnosis

- Diagnosis of weaponized disease in Dr. and E.R. can be turned on or off
- Diagnosis occurs if agent goes to Dr. office or E.R.
- Diagnosis can be correct or not – both type 1 and 2 errors
- Diagnosis results in
  - If at Dr. office – treatment or order test
  - If at E.R. – treatment, test or admission to hospital
- Symptoms vary in whether they are visible or require a low or high cost test
- Diagnosis is done via inference
  - The inference model was based on the Columbia QMR model which uses evoking strengths to infer likelihood of various disease
  - Differential diagnosis is possible corresponding to the onset symptoms
  - Can handle agents with multiple diseases
  - Each symptom has an evoking strength, P(D|S) (where D=disease, S=symptom)

Module 3: Tests for Diagnoses

- Diagnostic tests vary in
  - Cost*
  - Time to get a result
    - 3 categories
      - Immediate (visible)
      - Simple (timing)
      - Complex (timing)
  - Results from test impact

Module 3: Diagnostic Latencies

- Dr. and ER diagnoses take a while to send reports
- Dr. report latency is based on:
  - Source of data
  - Chart
- ER report latency is based on:
  - Source of data
  - Chart

Module 4: Agent Behavioral Model

- Agents
  - Roles – father, teacher …
  - Socio-demographic economic status
  - Location (longitude, latitude coordinates*)
  - Behaviors
    - Interact – communicate, be exposed, be infected, infect
    - Recreate, school, work
    - Seek treatment – OTC, Dr. E.R. (based on self diagnosis)
    - Get medical info – phone, web
    - Move (natural mobility)
  - Ego net – a network of relationships centered around self
  - Natural biological time, e.g., sleeping for 8 hours a day – every 4 hr or by day output
  - Mental model of the disease

Module 4: Agent Interaction & Knowledge

- *Build interaction graph*:
  - For each agent that can interact, choose a random agent *A* from the agent's ego net
  - Compute the probabilities of interaction with *A* due to common knowledge P(K|I) and proximity P(D|I)
  - Compute the probability of interaction P(I) as a weighted combination of P(K|I) and P(D|I):

    $P(I) := W_{,spatial} * P(K|I) + (1 – W_{,spatial})*P(D|I)$
  - Throw a random die *R* : if *R* < P(I), then add *A* to the agent's partner list
- *Compute interaction effects*:
  - For each agent, determine if the agent exchanges knowledge with each of its partners
  - Update agent interaction timing info (used to determine if agent should interact each tick)

Module 4: Disease Exchange

- *Disease exchanges*:
  - For each partner of an agent (computed during the interaction step), throw a random die against the transmissivity of each communicable disease affecting the agent, modulated by susceptibility
  - If the die roll fails and the partner does not already have the same disease instance (*strain*), infect the partner with the disease
  - Do the same check for the agent for each communicable disease infecting the partner

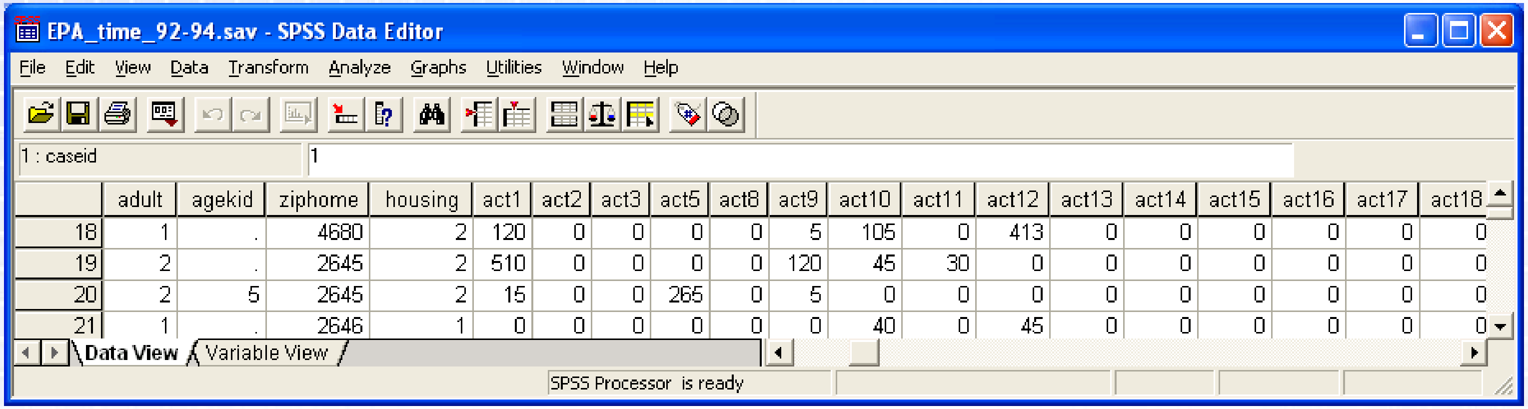
Module 4: Entertainment

- Agents may spend time on recreational activities
- Preferred entertainment is a function of
  - Agent demographics
  - Time of week/day
  - Normal versus holiday/school vacation days
  - Current health (varies by severity of illness)
- Entertainment types
  - External (go to shopping mall, sports event, concert, restaurant)
  - Home-based (read, watch TV, chat)
- Type of entertainment can affect likelihood of
  - Being an attack victim
  - Having knowledge about an attack

Module 4: Agents and Entertainment

- Two step determination process:
  1. Does the agent recreate this tick?
  2. Where does recreation take place?
- Probability based, using time-use survey data:
  - Season of year
  - Generation (child of 18 or less versus adult)
  - Gender
  - Day of week
- Additional enhancements:
  - Time of day
  - Holidays (using the school_calendar package)

Module 4: Probability of Recreation



	adult	agekid	ziphome	housing	act1	act2	act3	act5	act8	act9	act10	act11	act12	act13	act14	act15	act16	act17	act18
18	1	.	4680	2	120	0	0	0	0	5	105	0	413	0	0	0	0	0	0
19	2	.	2645	2	510	0	0	0	0	120	45	30	0	0	0	0	0	0	0
20	2	5	2645	2	15	0	0	265	0	5	0	0	0	0	0	0	0	0	0
21	1	.	2646	1	0	0	0	0	0	0	40	0	45	0	0	0	0	0	0

T(leisure) = avg((ACT23..25+ACT28+ACT30..31+ACT61..99)/(24*60))

Spring (3/21-6/20) Leisure as a Proportion of the Day from the EPA Time Use Survey 1992-94

Leisure Proportion of Day

Mean

Generation	GENDER OF RESPONDENT	DAY OF THE WEEK THE DIARY REFERS TO SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
Child	FEMALE	.38753655	.25492063	.27854610	.29415954	.23860480	.29462963	.37991898
	MALE	.40271868	.26675347	.25818452	.22735566	.26339286	.33011364	.41867766
Adult	FEMALE	.36795546	.26096347	.25282818	.23604798	.23675259	.25868056	.31142757
	MALE	.38601876	.21235450	.22445437	.21614583	.22740784	.23852778	.32753923

Module 4: Leisure EPA 92-94 time use

Probability of Recreation – Spring, Normal Day							
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Female Child	**0.4024**	**0.2862**	**0.2715**	**0.2737**	**0.2716**	**0.3077**	**0.3921**
Male Child	**0.4001**	**0.2859**	**0.2668**	**0.2593**	**0.2472**	**0.3142**	**0.4112**
Female Adult	**0.3603**	**0.2656**	**0.2438**	**0.2401**	**0.2457**	**0.2599**	**0.3264**
Male Adult	**0.3892**	**0.2454**	**0.2359**	**0.2304**	**0.2410**	**0.2411**	**0.3318**

Module 4: Validation and Tuning of Entertainment

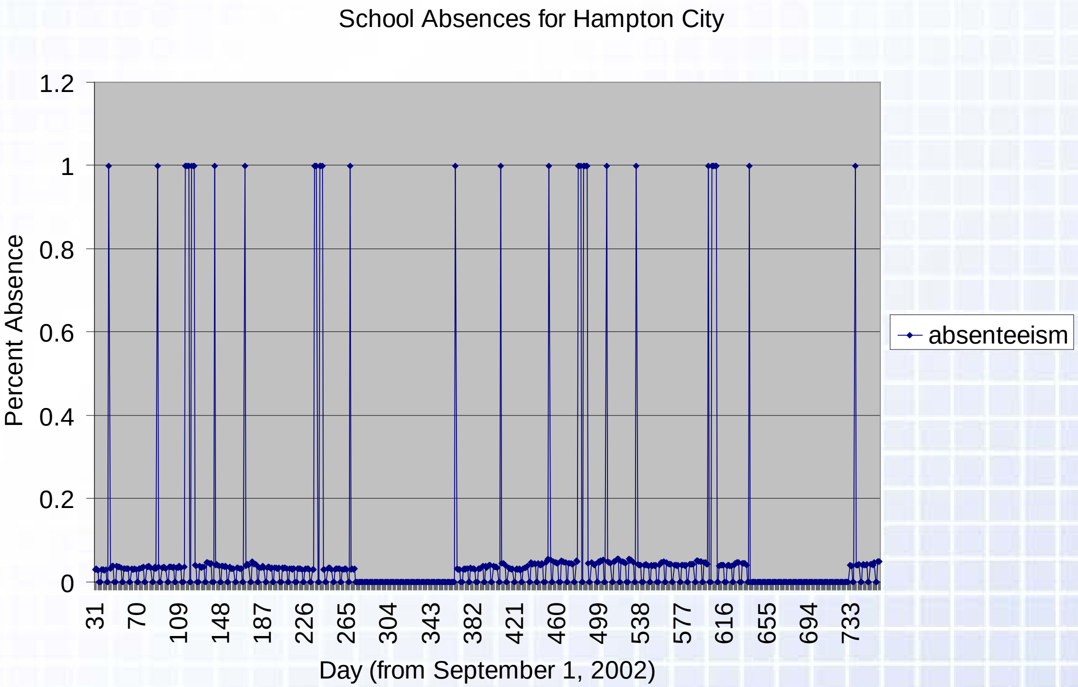
- Primary data source: EPA Time Use survey (1994) – access to raw data
- Two critical data types:
  - Time spent by activity category
  - Time spent in locations
- Must infer certain critical values:
  - Time spent in recreation at specific locations
  - Holidays
  - Time of day variations
- Validated
  - Annual recreation rates
  - School absenteeism rates

Module 4: School Absenteeism

- Absenteeism occurs due to
  - Illness
  - Skipping
  - Other
- Probability of non illness absence set by school level
- Data from NCES Indicator 17 & Indicator 42-1
- Data from Veridian
- Non illness absence determined randomly
- Minor exceptions
  - Higher absenteeism prior to and after weekend holiday
  - No school on weekend, summer, holidays

Module 4: Behavioral Report - School Absenteeism

- Standard
  - School id
  - Tick
  - Report tick
  - Registered
  - Absent
- Reports are always in morning, 3 tick delay
- No school on weekends, summer
- Possible info that can be recorded
  - Home zip code of absent student
  - Characteristics of absent student



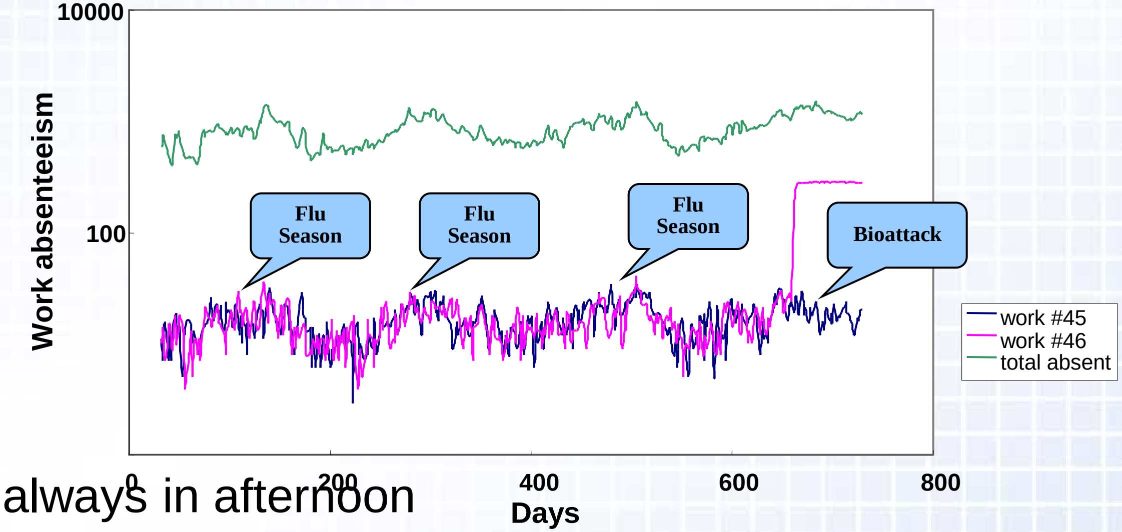
School Absences for Hampton

Module 4: Work Site Behavior

- Absenteeism occurs due to
  - Illness
  - Other
- Data from Veridian
- Non illness absence determined randomly at pre-specified level
- Minor exception
  - Higher absenteeism prior to and after weekend holiday
- Phone calls
  - Data from IBM
- Web visits
  - Data from on-line hit rate for medical sites

Module 4: Behavioral Report - Work Site

- Standard
  - Workplace id
  - Tick
  - Report tick
  - Registered
  - Absent
  - Phone calls



- Work report – 3 tick delay, always in afternoon
- Work is 5 day work week, 2 ticks long, 12 months
- Possible info that can be reported
  - Home zip code of absent worker
  - Characteristics of worker

Module 4: Seeking Treatment

- Propensity to seek treatment affected by
  - Socio-demographic position (age, race, gender)
  - Socio-economic status
  - Severity of visible symptoms
  - Medical history
- Type of treatment also impacted by availability
  - E.g. can't go to pharmacy or Dr. if closed
- Reporting delays
  - Based on SME estimates and IBM data

Module 4: Agent Self-diagnosis

- *Do self-diagnosis*:
  - For each agent, for visible symptoms compute the total symptom severity *S* of diseases affecting the agent
  - Check *S* against user-specified thresholds to determine agent behavior:

$S < T_{,pharm} \rightarrow$ No change to *Default Next State*

$T_{,pharm} < S \leq T_{,clinic} \rightarrow$ Send agent to pharmacy on next tick

$T_{,clinic} < S \leq T_{,ER} \rightarrow$ Send agent to clinic on next tick

$S > T_{,ER} \rightarrow$ Send agent to ER on next tick

Module 4: Agent's Self Diagnosis cont.

If agent goes to pharmacy then symptoms determine purchase with some probability. Illustrative table.

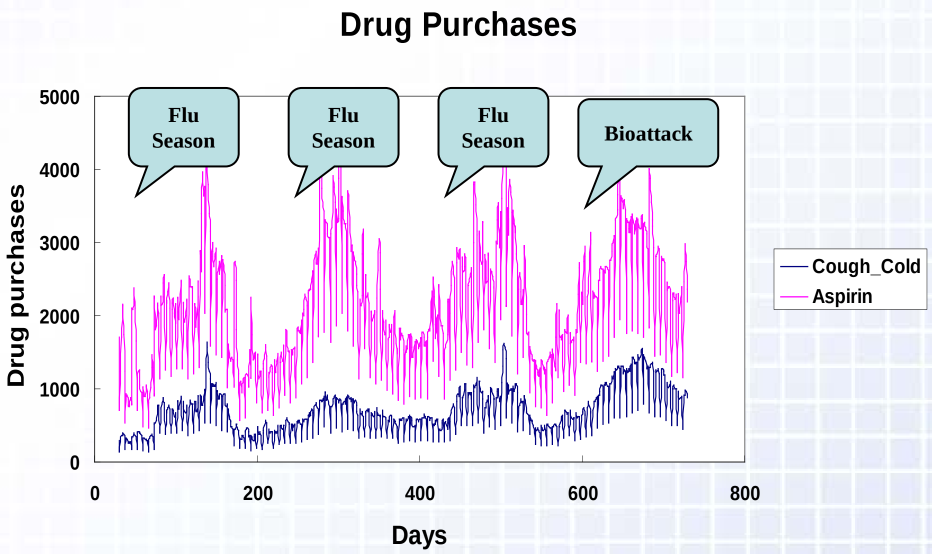
	Coughing	Sneezing	Muscle pain	Fever	Headache	Diarrhea
Cough medicine	☛					
Cold medicine		☛				
Cold+cough medicine	☛	☛				
Cold,cough, fever medicine	☛	☛		☛		
Analgesic			☛	☛	☛	
Anti-diarrheal						☛

Module 4: Pharmacy Behavior

- Agents go to the pharmacy nearest work if at work or nearest home if at home
- Children under 12 not allowed to purchase
- Planned:
  - Purchasing for others
  - Variation in purchased amount, multiple purchases
  - Purchasing increases in December

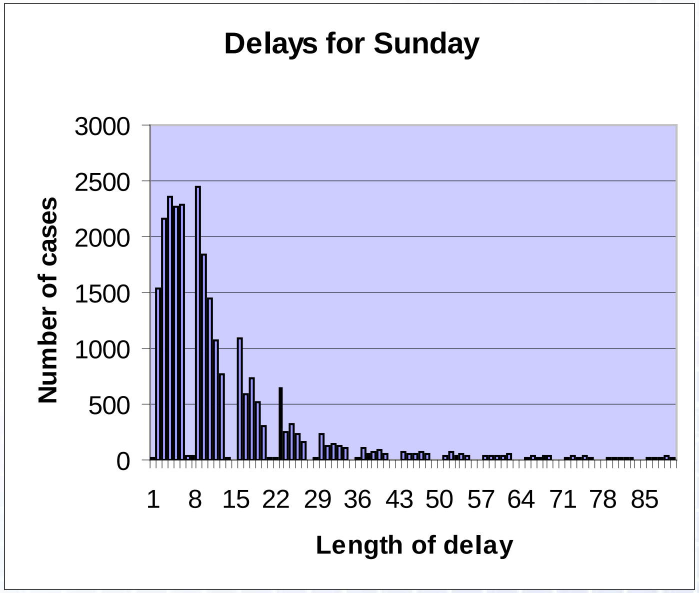
Module 4: Behavioral Report Over the Counter Purchases

- Standard
  - Pharmacy id
  - Tick
  - Report tick
  - Number of purchases of
    - Cold
    - Cold-cough
    - Cough
    - Analgesic
    - Anti-diarrheal
    - Kleenex
    - Orange Juice
- Reporting delay 3 ticks
- Open 7 days a week, reduced Sunday hours



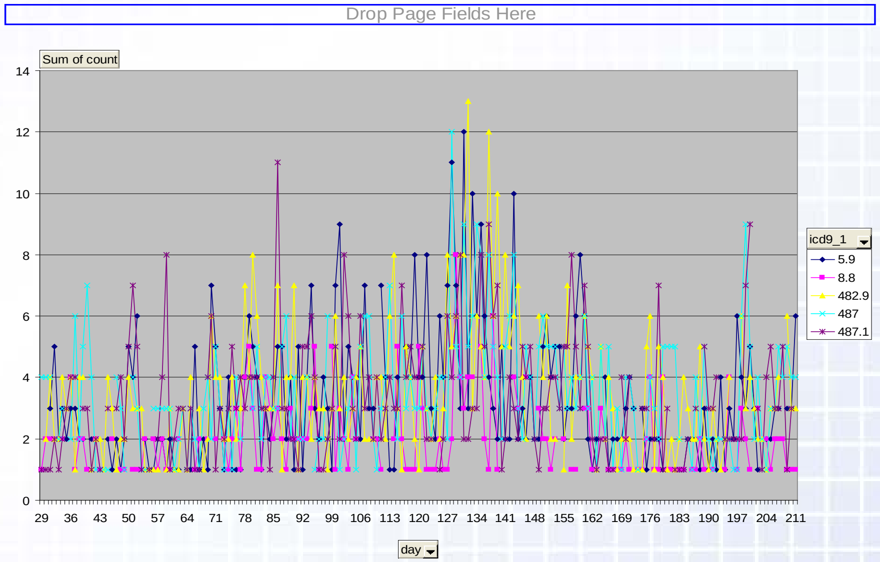
Module 4: Behavioral Report on Insurance Claim Reports

- Standard
  - Tick
  - Report tick
  - Call tick
  - Icd9 of disease
  - Icd9 of 3 major symptoms
  - Doctor id
  - Patient
    - Home zip code
    - Work zip code
    - Age
    - Gender
  - Doctor zip code
- Reporting delay – varies by day of week, range 0 to greater than 90 days, based on empirical data



Module 4: Behavioral Report on Emergency Room Registration

- Standard
  - Hospital id
  - Tick
  - Report tick
  - Icd9 of disease
  - Icd9 of top 3 symptoms
  - Patient
    - home zip code
    - work zip code
    - Patient age
    - Patient gender
    - Previous zip code
    - Disposition
    - Disposition tick
- Reporting delay, 3 ticks
- Higher utilization at night, weekends, holidays



Module 5: Geometry

- Conversion between
  - ZCTA – UTM – Lat/Lon
- Linkage of ZCTA to USPS
- All agents/locations have location
- Assorted pre and post-processors
- Location impacts choice of E.R., Dr. Pharmacy …

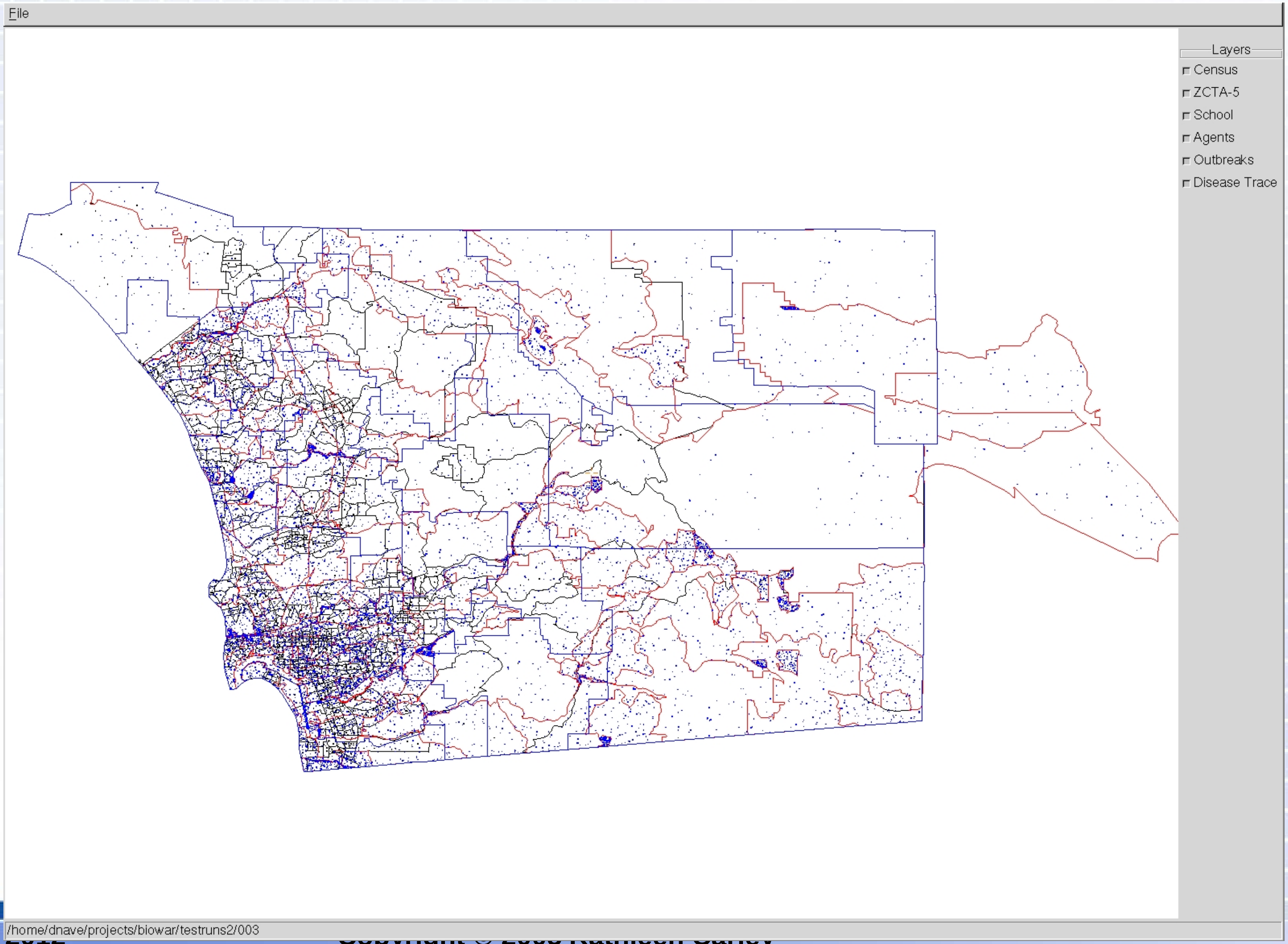
Geometry In BioWar

- Census cartographic boundaries enclose simulation area – polygon vertices specified by longitude/latitude
- Positions of locations, agents, outbreaks, and attacks specified by longitude/latitude
- Most distances computed in longitude/latitude
- Outdoor attack position, agent positions dynamically converted into UTM coordinates to compute distances from attack
- Geometry composed of several simple classes

Geometry In BioWar (Cont'd) – Objects

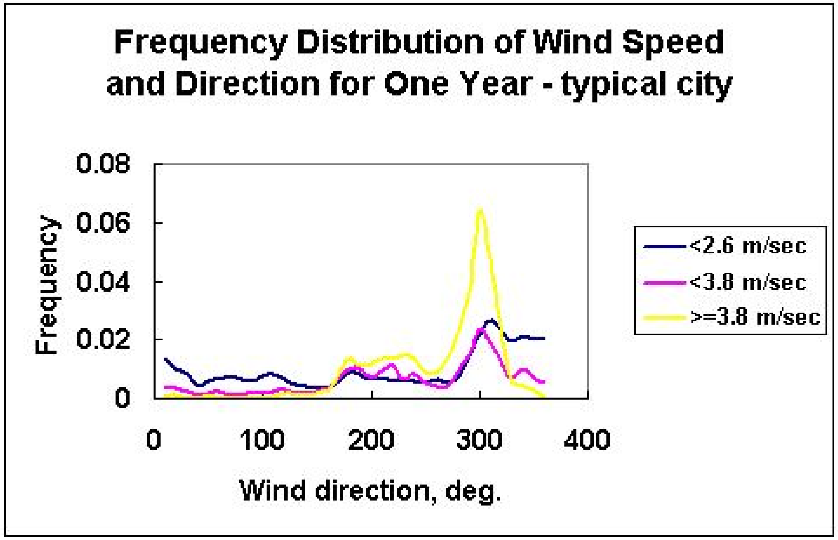
- Points
  - Longitude/latitude and UTM coordinate systems
  - Conversion between systems using several ellipsoid definitions
- Polygons
  - Generically programmed vertex positions (can use longitude/latitude, UTM, etc.)
  - Several geometric operations (e.g. point containment)
- Census tracts
  - Polygons with Census-assigned attributes
  - Currently handles school districts, ZCTA's, block groups, and census tracts
- Integration with Census Tiger/Line to improve e.g. agent distribution over census tracts

Example Area



Module 6: Weather - Wind Model

- Closely represents real meteorological conditions of city area taken from National Weather Service station observations.
- Assumes uniform values of wind speed and direction over the simulated area



Module 6: What varies for wind and what impacts it

Wind characteristics:

- Wind direction is the direction from which the wind comes
- Speed

Meteorology impacts wind

- Pasquill atmospheric stability class
- Temperature
- Mixing height

Current Wind Model assumes moderate insulation and thinly overcast cloud conditions.

Module 6: Sources of Data for Wind Model

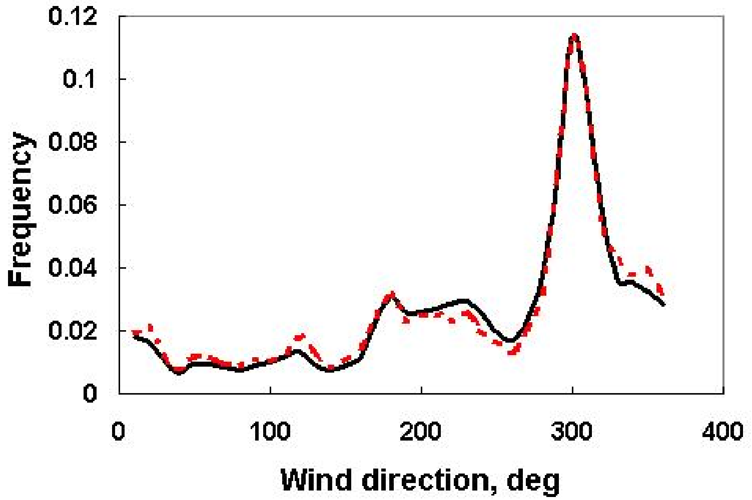
- Empirical data – www.epa.gov/scram001/
- Rod Barratt "Atmospheric Dispersion Modeling"
- D. Bruce Turner "Workbook of Atmospheric Dispersion Estimates"
- Meselson, Matthew "Note Regarding Source Strength", ASA Newsletter, article 01-6a (www.asanltr.com).

Module 6: Validation of Wind Model

- Validation was performed by comparing simulated wind data with the empirical data published at www.epa.gov/scram001/.

Wind Direction Frequency Distribution

for San Diego, CA



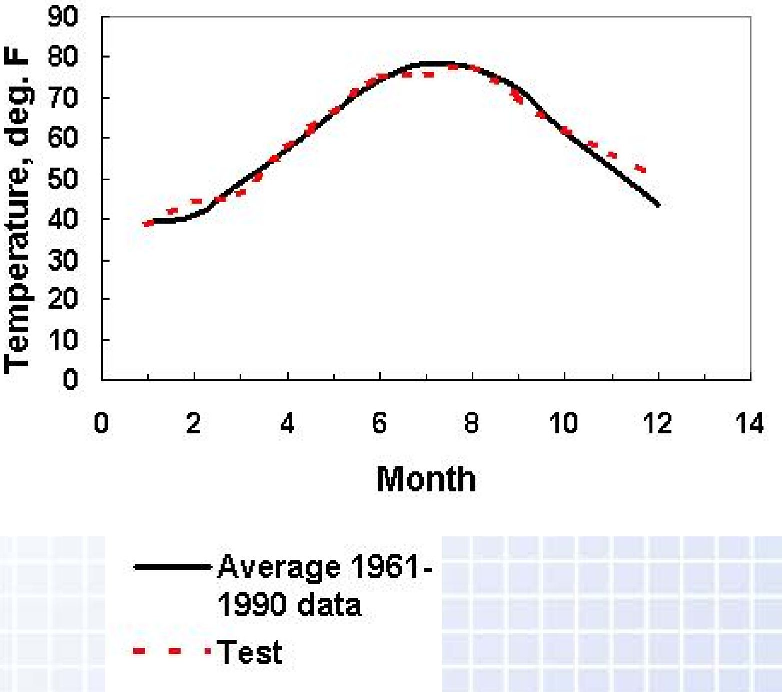
Black line – average 1990 – 1992 data

Red line – simulated data

Module 6: Weather - Climate Model

- Generates climate parameters - temperature, atmospheric pressure and precipitation
- Closely matches empirical data
- Climate parameters do not show local variations over the simulated region
- Source of the data – www.epa.gov/scram001/
- Validation was performed by comparing with historical data http://weather.gov/climatex.html

Norfolk average monthly temperatures



Module 7: Dispersion Module for Aerosolized Attack

- Inputs:
  - Emission information - location of the source and height of the release;
  - Meteorological parameters – Pasquill stability class, wind direction, wind speed
- Outputs:

  Dosage inhaled by the agent

- Module uses modified Gaussian Puff Equation to estimate total dosage from finite release.
- Described in geographical coordinate system (lat/lon) which is transformed to/from UTM coordinate system and local coordinate system with the origin at the bio-release point.

Module 7: Total inhaled dosage

The total dosage at a receptor at x, y, z from a finite release can be expressed as

$$\text{Dose} = [QB][\pi u \sigma_y \sigma_z]^{-1} \exp[-(1/2)(y/\sigma_y)^2] \exp[-(1/2)(H/\sigma_z)^2]$$

Source strength = Q spores

Breathing rate = B = $5 * 10^{-4}$ m^3/sec

Wind speed = u m/sec

Release height = H

Downwind (x), crosswind (y) distances and height (H) are in meters.

Meselson, Matthew “Note Regarding Source Strength”, ASA Newsletter, article 01-6a (www.asanltr.com).

Module 7: Dispersion Model Limitations

- Releases are assumed to be low-level.
- Deposition is negligible.
- Infectivity is independent from the puff travel time.
- The meteorological conditions are assumed to persist unchanged over the wind puff travel time from source to receptor

Module 7: Validation in terms of anthrax dispersion

- Wind speed = 5 m/sec
- Source strength = 0.01 g
- Pasquill atmospheric stability class “D”

Centerline Dose (spores) From Four Models

Distance	BioWar*	BioWar**	Meselson	Point V	TNO
1 km	29	166	106	317	281
2 km	9	55	36	109	91

*** Using Briggs urban conditions formulae**

**** Using Briggs open-country conditions formulae**

Module 7: Source of data for validation

- Meselson, Matthew “Note Regarding Source Strength”, ASA Newsletter, article 01-6a
- POINT V – “Methodology for Chemical Hazard Prediction”, DOD, 1980, p.17
- TNO – TNO Defense Research, Rijswijk, The Netherlands

- Possible reasons for the discrepancy:
  - BioWar uses the Briggs dispersion parameters formulae for urban conditions while sources above uses formulae for open-country conditions
  - Military methodologies tend to overestimate the effect in order to protect troops

Module 8: Attack Scenario Module

- Attacks are created following the scenarios
- Attack scenario allows maximum flexibility
- Attacks vary based on
  - Locations
  - Inside or outside of building
  - Date
  - Time of Day
  - Agent
    - Carrier – Airborne (contagion and non-contagion module), waterborne, food borne, other
      - Non-airborne not done
    - Severity
    - Pathogen (weaponized disease)

Module 8: Attack Parameters

- Land or airborne attack type
- Spray or explosion type (by selecting release efficiency)
- Specification:
  - Pathogen
  - Biomaterial mass, release height and efficiency
  - Random or fixed time/date
  - Random or fixed locations
  - Single point or multi-point
  - Impact (low, medium, high based on number of people actually affected)

Module 8: Attack Scenario Examples

- Generate a medium, single-point spray attack between 100 and 200 ticks at an altitude of 20m, using 1.25kg of material for an attack at 5% efficiency.

```
out medium anthrax_inhalational 100 200 1.25kg .05 20m ;
```

- Generate a large, multi-point airborne attack at 22:00 on July 4, 2002 an altitude of 300m, using 25kg of material for an attack at 10% efficiency. Distribute 7 bombs along an attack line of 1.5 km

```
out large anthrax_inhalational 2002/7/4 22:00 25kg .1 300m
  1.5km 7 ;
```

Module 9: What data post processors are available

- Postprocessors perform output data transformation to the format required by the customer
  - Create “corner” files, for file integrity check
  - Collapse output files from “by tick” to “by day” representation and insert “0” values for display
  - Extract EPI (epidemiological curve) data for any simulated disease

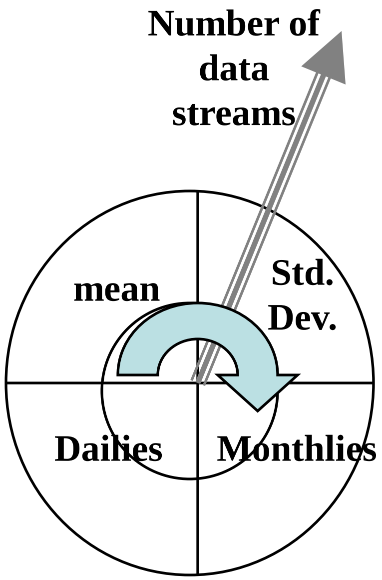
Module 10: Verification & Validation

- Internal Tuning
  - Existing data sets to parameterize
    - Reporting delays
    - Disease profiles
    - Agent social networks
    - Age, race, gender, economic differences on behavior and susceptibility
    - Variation in behavior by time of day, day of week, month, season
    - Usage of IT
  - Sources
    - Behavioral surveys
    - Nursing studies
    - CDC reports
    - Communication studies
    - OTC purchases
  - City profiling
    - Census data
    - School district
    - Maps

- Validation – emergent behavior compared to real data
    - Death reports
  - General behavior
    - Disease replication for historic cases
    - Pharmacy purchases
    - Cold shelf and influenza spike
  - Influenza
    - Grade School Absenteeism
    - ER reports
    - OTC purchases
  - Level
    - General pattern
    - Mean, std
    - Variation in disease reports by day of week, month, season, local

Module 10: What Data Streams is Validation Done On

Data Stream	C2	C3
Work absenteeism	**Yes**	**Yes**
School absenteeism	**No**	**Yes**
ER visits	**Yes**	**Yes**
Doctor visits	**Yes**	**Yes**
OTC drug purchase	**No**	**Yes**
Sentinel trace	**No**	**No**
Network distribution	**No**	**Yes**



Module 10: What validation or tuning has been done

- Work absenteeism within the lower & higher empirical bounds
- School absenteeism within the lower & higher empirical bounds
- Doctor visits within the lower & higher empirical bounds
- ER visits within the lower & higher empirical bounds
- Drug sales per group is near the empirical mean
- Face validation of a sentinel population trace
- Automated output check

Module 10: Sources of Data for Validation

- NCES Indicator 17 & Indicator 42-1, for calculating school absenteeism
- CDC Advance Data, from Vital and Health Statistics, no. 326, 2002, for calculating ER visits
- CDC Advance Data, from Vital and Health Statistics, no. 328, 2002, for calculating doctor visits
- 1997 US Employee Absences by Industry Ranked (http://publicpurpose.com/lm-97absr.htm) for determining work absenteeism
- OTC Sales by Category from AC Nielsen (http://www.chpa-info.org/statistics/otc_sales_by_category.asp) and PSC's FRED data for pharmacy OTC drug sales

Module 10: Empirical School Absenteeism Bounds

- Data from NCES Indicator 17 & Indicator 42-1
- NCES Indicator 42-1 gives total absenteeism rate of 4.9% for 8th graders in urban fringe/large town
- NCES Indicator 17 gives the absenteeism reasons of illness of 53.1%, skipping 9.0%, others 37.9%.
- For 10th graders, the corresponding total absenteeism rate is 6.2%, absenteeism due to illness of 45.4%, skipping 15.6%, others 39.0%
- For 12th graders, the corresponding total absenteeism rate is 8.6%, portion of it due to illness is 34.2%, skipping 26.1%, others 39.7%
- As we don't have reasons other than illness or skipping in C3, the lower bound for all schools is 3.04%, with the upper bound of 5.18% absenteeism rate

Module 10: School Absenteeism

City, percent of simulated population	Empirical lower bound	Empirical higher bound	No Attack (mean)	Anthrax (mean)	Smallpox (mean)
Norfolk, 20%	3.04%	5.18%	3.45%	3.75%	3.55%
Pittsburgh, 20%	3.04%	5.18%	3.52%	4.67%	4.46%
San Diego, 20%	3.04%	5.18%	3.78%	3.81%	5.57%
Veridian Norfolk, 20%	3.04%	5.18%	3.73%	4.05%	4.31%

Module 10: Empirical Work Absenteeism Bound

- Data from the 1997 US Employee Absences by Industry Ranked
- As we don't yet have the specifics of workplace types in C3, we take the lower bound to be the lowest absence rate of any industry type, the higher bound to be the highest.
- So, from the data, we have the lower bound of 2.3% and the higher bound of 4.7% absenteeism rate.

Module 10: Work Absenteeism

City, percent of simulated population	Empirical lower bound	Empirical higher bound	No Attack (mean)	Anthrax (mean)	Smallpox (mean)
Norfolk, 20%	2.30%	4.79%	2.72%	4.65%	2.82%
Pittsburgh, 20%	2.30%	4.79%	2.77%	5.79%	3.99%
San Diego, 20%	2.30%	4.79%	3.26%	4.99%	5.78%
Veridian Norfolk, 20%	2.30%	4.79%	3.16%	5.50%	3.81%

Empirical Doctor Visits Bound

- Data from CDC Advance Data, Vital & Health Statistics, No. 328, 2002
- Table 1 of the report shows MSAs (metropolitan areas) have 294.6 visits per 100 persons per year
- The lower bound is based on major disease categories, while the higher bound is based on all disease categories in the simulation
- Table 11 of the report gives 14.1% of all the causes of visits to fall within major disease categories of infectious & respiratory diseases, and 54.7% for all disease categories in the simulation
- This gives us the lower bound of 0.415 visits per person per year and the higher bound of 1.611 visits per person per year

Module 10: Doctor Visit (visit per person per year)

City, percent of simulated population	Empirical lower bound	Empirical higher bound	No Attack (mean)	Anthrax (mean)	Smallpox (mean)
Norfolk, 20%	0.415	1.611	0.499	0.476	0.499
Pittsburgh, 20%	0.415	1.611	0.493	0.485	0.573
San Diego, 20%	0.415	1.611	0.726	0.753	0.796
Veridian Norfolk, 20%	0.415	1.611	0.707	0.821	0.738

Module 10: Empirical ER Visits Bound

- Data from CDC Advance Data, Vital & Health Statistics, No. 326, 2002
- Table 1 of the report shows MSAs have 37.6 visits per 100 persons per year
- The lower bound is based on major disease categories, the higher bound on all disease categories in the simulation
- Table 7 in the report gives us 14.8% of all causes tp fall within major disease categories of infectious & respiratory illness, and 77.7% of all disease categories of the 62 disease present in the simulation
- So the lower bound is 0.056 visits per person per year, the higher bound 0.232 visits per person per year

Module 10: ER Visit (visit per person per year)

City, percent of simulated population	Empirical lower bound	Empirical higher bound	No Attack (mean)	Anthrax (mean)	Smallpox (mean)
Norfolk, 20%	0.056	0.232	0.112	0.108	0.112
Pittsburgh, 20%	0.056	0.232	0.109	0.106	0.129
San Diego, 20%	0.056	0.232	0.149	0.159	0.188
Veridian Norfolk, 20%	0.056	0.232	0.161	0.187	0.168

The Need for Validation Automation

- Validation is difficult to do manually due to model complexity - the significant number of input and model parameters, output variables
- Scaling BioWar up to take in more models – local models and diverse secondary data streams – would increase the code size
- Real-time revalidation of BioWar to changing real world situations is of importance
- An automated tool that analyzes software and rates its reliability by examining the response surface relative to empirical data is needed

Wizer: Automated Validation

- Response surface methodology:
  - collection of mathematical and statistical techniques (e.g., gradient descent search) for the modeling & analysis of problems in which a response of interest is influenced by several variables and the objective this response.
  - BioWar has a complex response surface.
- Putting BioWar in Spec can be viewed as a multi-dimensional numeric & symbolic optimization problem
  - E.g., school absenteeism is influenced by student health status, skipping, or other reasons such as school district announcements.
  - Within these symbolic variables, there are numeric values to denote the probability, the trends, etc.
- Wizer (What-If AnalyZER)
  - extends response surface methodology by performing knowledge-intensive search via an inference engine with the search is in form of both knowledge inferences and simulation “virtual experiments”
  - instead of doing conventional mathematical & statistical calculations
- Better faster validation and tuning

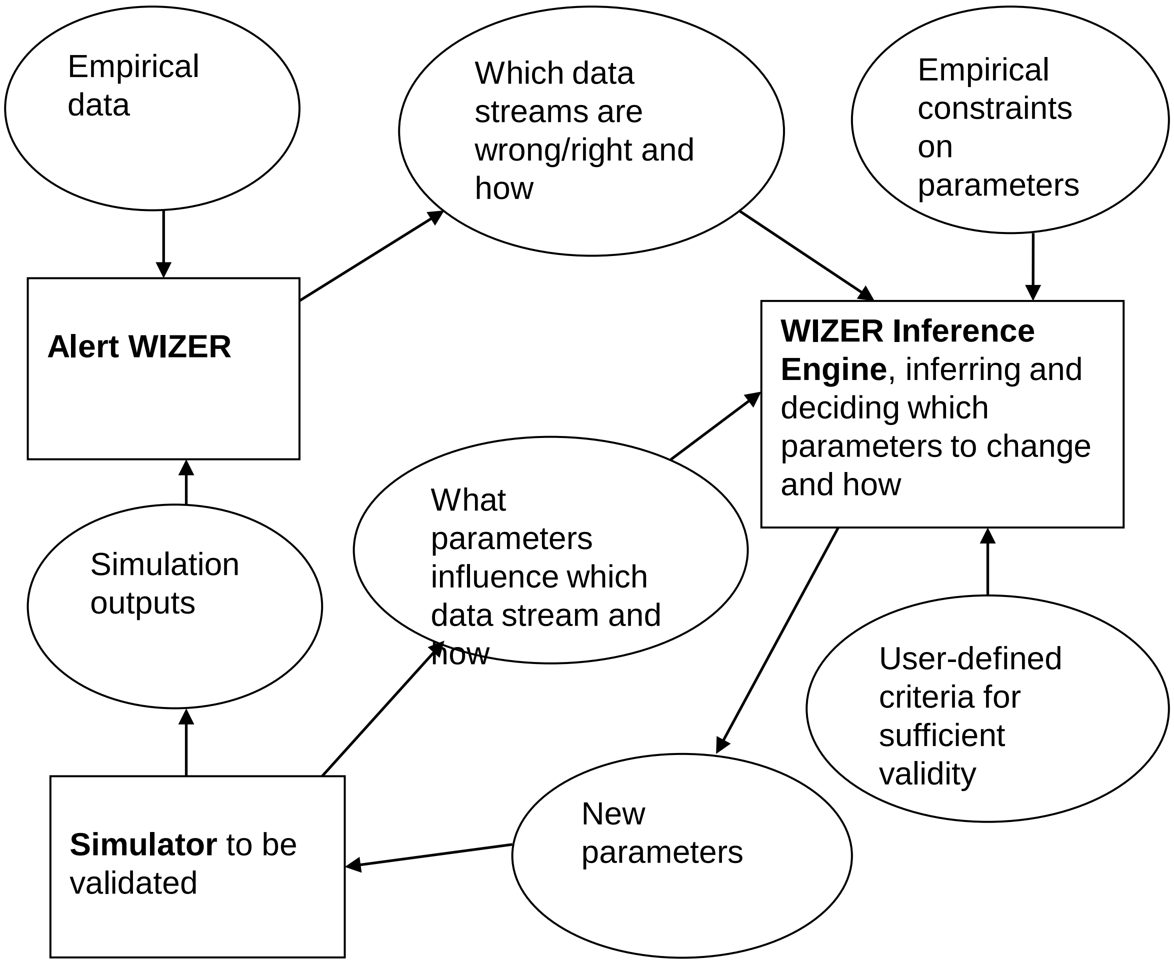
Wizer: Definition

- Wizer is
  - a tightly-coupled inference and simulation engine
  - that extends the response surface methodology
  - to deal with high dimensional, symbolic, stochastic, emergent, and dynamic nature of complex multi-agent systems
  - by performing knowledge-intensive data-driven search steps via an inference engine constrained by simulation
  - and by explaining the reasoning behind inferences using both the simulation and the inference engine

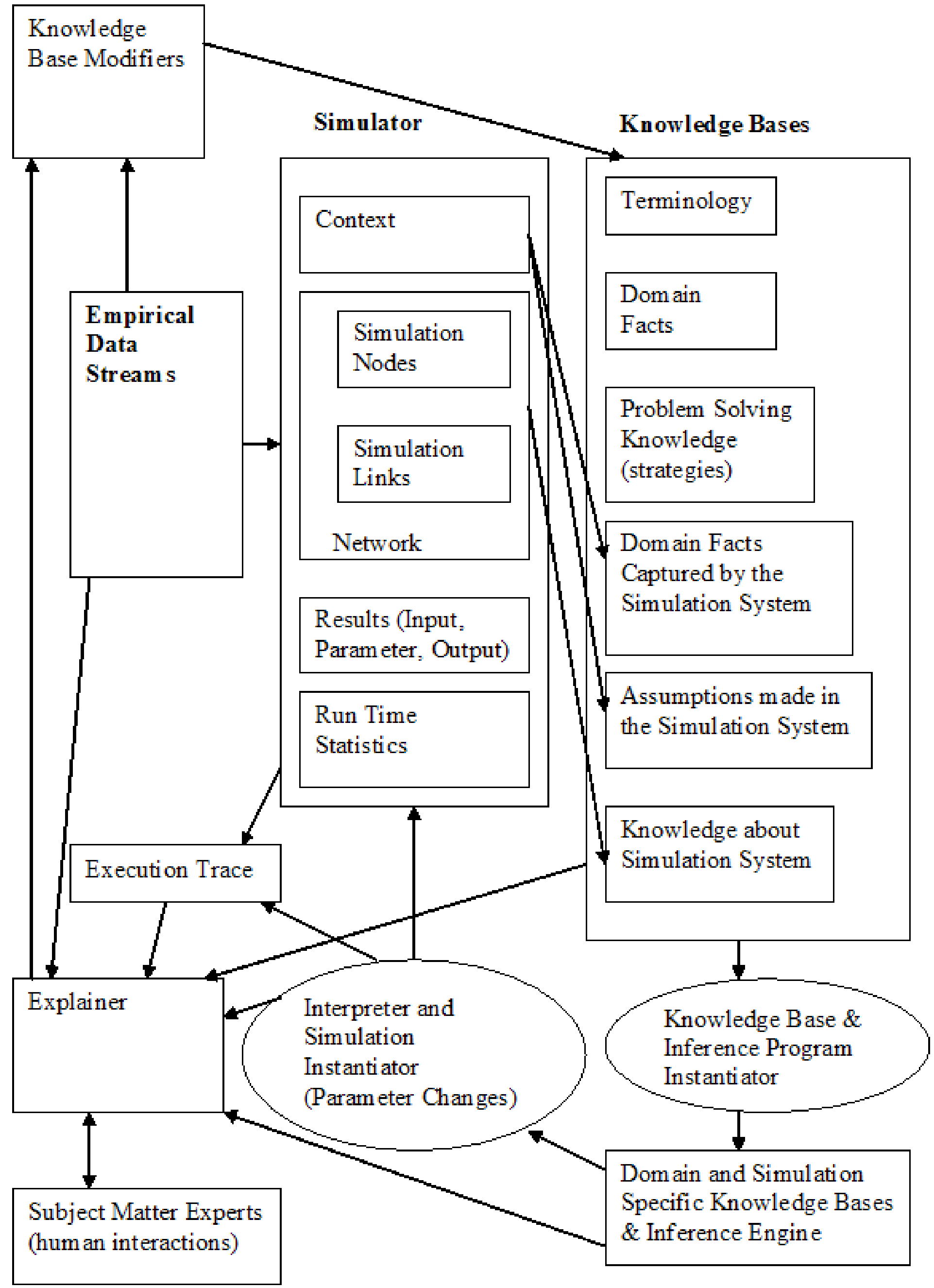
Wizer: Version 0 Implementation

- *[Wizer version 0 or "Alert Wizer", implemented & deployed:]* Takes the simulated output data and the set of validation specs and "sets off an alarm" if
  - When the distribution is known – if the simulated data is ever more than 1 std-dev away from from the spec
  - When the distribution is not known – if the simulated data is outside the allowable range on the spec
- Automated search for set of changes to move the simulation back within spec. To do this, Wizer utilizes social, epidemiological, geographical, etc. knowledge via inference engine. Wizer can be viewed as an intelligent search step generator for simulation. The inference engine and simulation components are tightly-coupled in Wizer.
- Note: user is able to specify to this system how mutable the input parameters are
  - for some parameters you can vary over a wide range while other parameters are fixed
  - how mutable would depend on the quality of the data underlying them

Wizer Diagram



Tightly-coupled Simulation and Inference Engine Components of Wizer



Acknowledgements