STEM: The Spatiotemporal Epidemiological Modeler

http://www.eclipse.org/stem/

M. Davis, D. Doerr, S. Edlund,
K. Hu, J. Kaufman, S. Renly

jhkauf@us.ibm.com

August 2012

Outline

- The STEM framework
- Research Examples
  - Seasonal Influenza with Israeli CDC
  - H1N1 with DGF Mexico
  - Vector borne diseases
    - Malaria
    - Dengue Fever
  - Oral Polio Vaccine (OPV)
  - Modeling food production and food borne disease in STEM
- Proactive modeling of food borne disease outbreaks

STEM is an open source *framework* for Epidemiological Modeling

STEM Website: http://www.eclipse.org/stem/

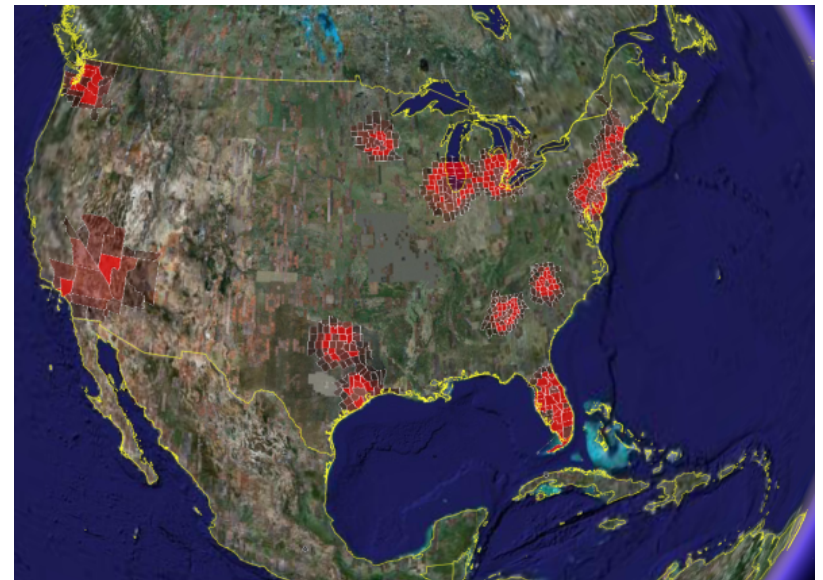STEM Wiki (Documentation): http://wiki.eclipse.org/index.php/STEM

STEM Newsgroup: http://www.eclipse.org/forums/index.php?t=thread&frm_id=72

STEM on Wikipedia: http://en.wikipedia.org/wiki/Spatiotemporal_Epidemiological_Modeler
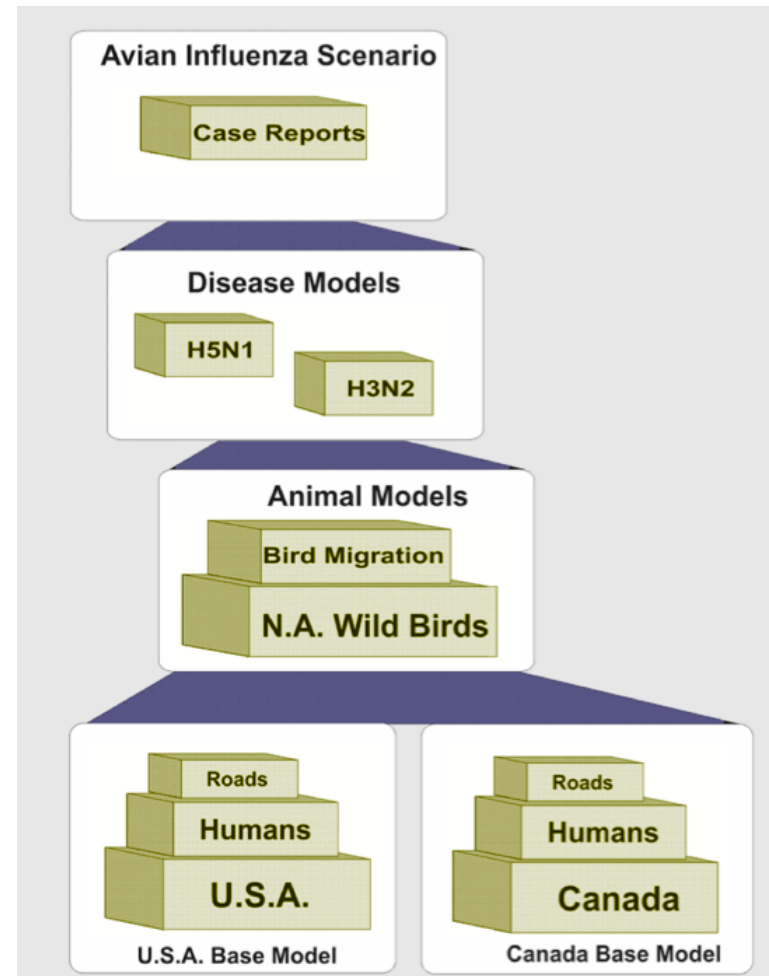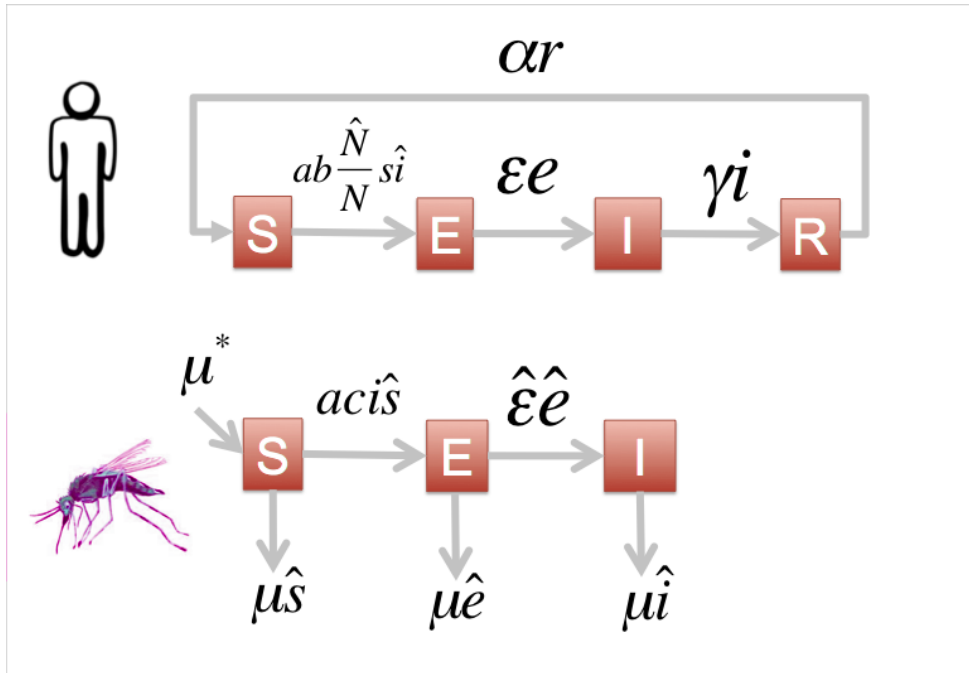
Video Tutorials on YouTube
http://www.youtube.com/watch?v=LfiibQX4IFE   *English*
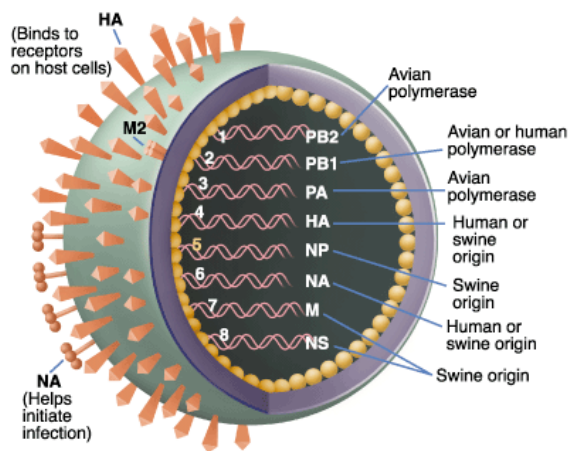http://www.youtube.com/watch?v=3S5DbjCHsx4   *Spanish*
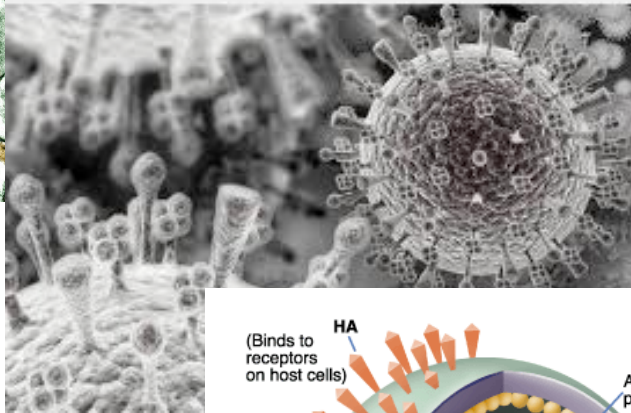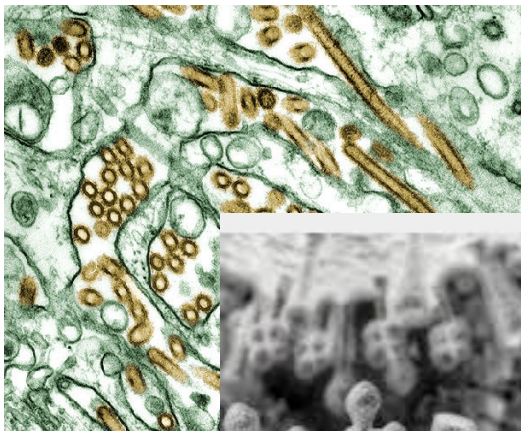
Why a Framework?
Modeling Disease involves building models on top of models

Why Open Source ?

The reservoir of emerging disease is the animal population

It takes a Community!!

Bundesinstitut für Risikobewertung
Epidemiologie und Zoonosen

University
of Vermont (UVM)

University of Pittsburgh

UPMC

STANFORD
UNIVERSITY

NTI
Working for a Safer World

MECIDS

Search for
Common Ground

LOCKHEED MARTIN

iCDC

U. Helsinki

Northrop
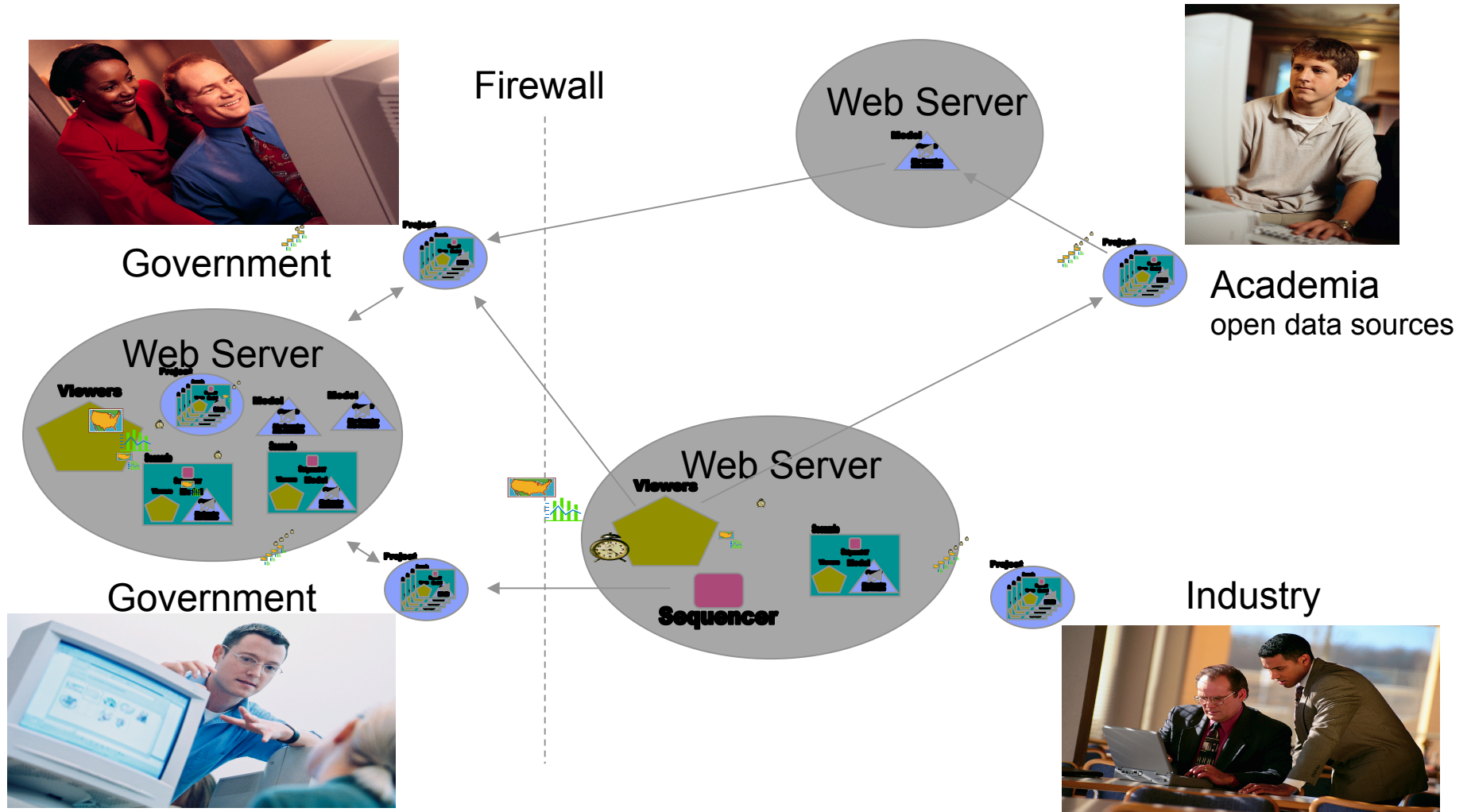Grumman

UC Davis

Massachusetts
Institute of
Technology

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Why Eclipse?
## Community and Collaboration Supported by the Eclipse Foundation and the *EPL*

Firewall

Web Server

Government

Academia
open data sources

Web Server

Viewers

Government

Web Server

Viewers

Sequencer

Industry

STEM uses an new *component software* technology

- OSGi
- All STEM components can be dynamically loaded as separate bundles or plug-ins:
    - Disease Models are Plug-ins
    - Population Models are Plug-ins
    - Datasets are Plug-ins
    - Viewers are Plug-ins
        - Map view
        - Google Earth
        - Google Maps
        - NASA World Wind, etc.

- Each plug-in can be independently developed, deployed, and used with declarative software extension points

- Users can
    - create their own scenarios for a country or region
    - Collaborate, *exchange*, models and scenarios
    - Easily combine, manage, and partition plug-ins across security domains
    - Build on existing models and create new ones, making STEM extensible, flexible, and re-usable

- Today STEM is an Eclipse Application.
    - In the future we wish to make STEM available also **as a service** using the new OSGi on Server Technology from Eclipse
    - Also exploring ways to make STEM easy to use by any SME
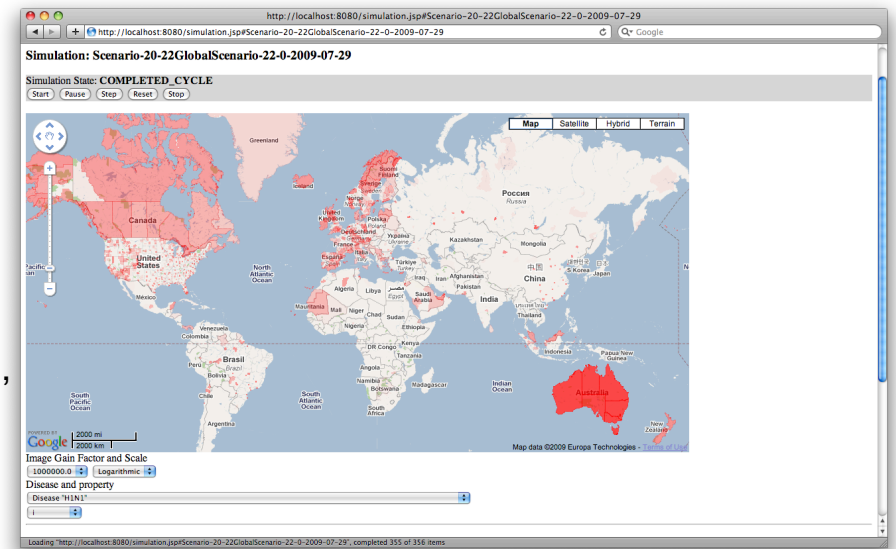
STEM treats the World as a "Graph"

Any problem that can be described as a graph can be studied with STEM
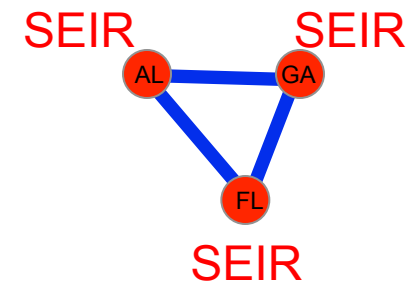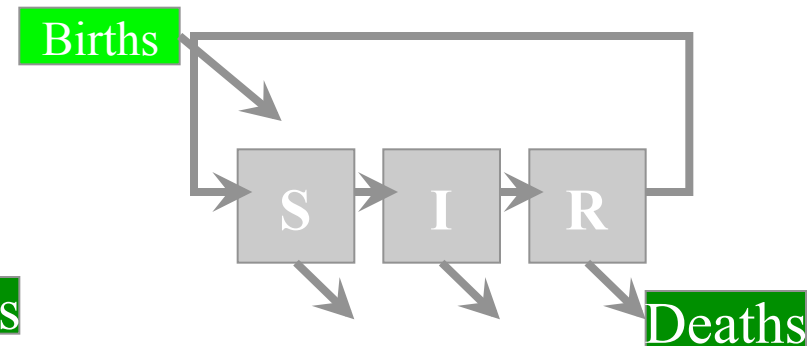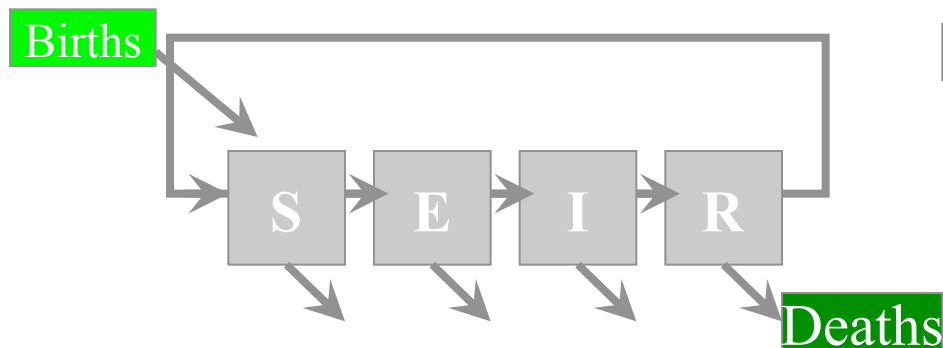

- "Graph"={Nodes, Edges; with labels & decorators}

- Graph Framework makes it easy to *build one model on top of another*

- Existing plugins define geography, transportation systems, populations, land area, etc. for the planet (244 countries and dependent areas)

STEM *solves differential equations*
Epidemiological Compartment Models

Standard SEIR model   Standard SIR model

S: susceptible
E: exposed but not yet infectious
I: infectious
R: recovered

STEM includes many compartment models

- Textbook models SI, SIR, SEIR and stochastic variants

- Advanced and state of the art models

  - Multiple population models (including demographics)
  - Zoonotic models
  - Vector models
  - Food production models
  - Foodborne disease models

All models are spatiotemporal

- Allows uses to run global models

- Allows users to construct custom regional models

  - Farms
  - Feeding facilities
 - Slaughterhouses

- Pluggable transportation models

$$\frac{dS_j(t)}{dt} = -\beta(t)S_j(t)\frac{\sum_{k=1}^{K}m_{jk}I_k(t)}{\sum_{k=1}^{K}m_{jk}P_k(t)} + \alpha R_j(t) + bP_j(t) - \mu S_j(t)$$

$$\frac{dI_j(t)}{dt} = \beta(t)S_j(t)\frac{\sum_{k=1}^{K}m_{jk}I_k(t)}{\sum_{k=1}^{K}m_{jk}P_k(t)} - \gamma I_j(t) - \mu I_j(t)$$

$$\frac{dR_j(t)}{dt} = \gamma I_j(t) - \alpha R_j(t) - \mu R_j(t)$$

Need to Explore a Large Phase Space of Parameters!!
Nelder-Mead Simplex Algorithm

Automated Experiments run many simulations, automatically walking the space of models parameters to find an optimal set of parameters for a given model.
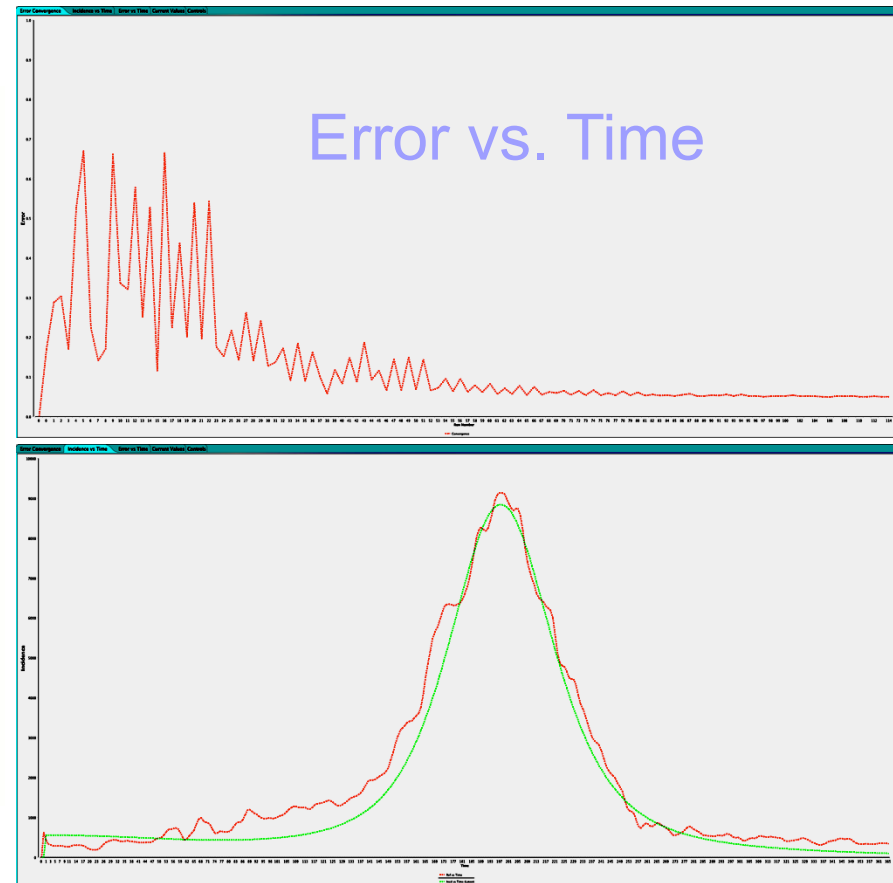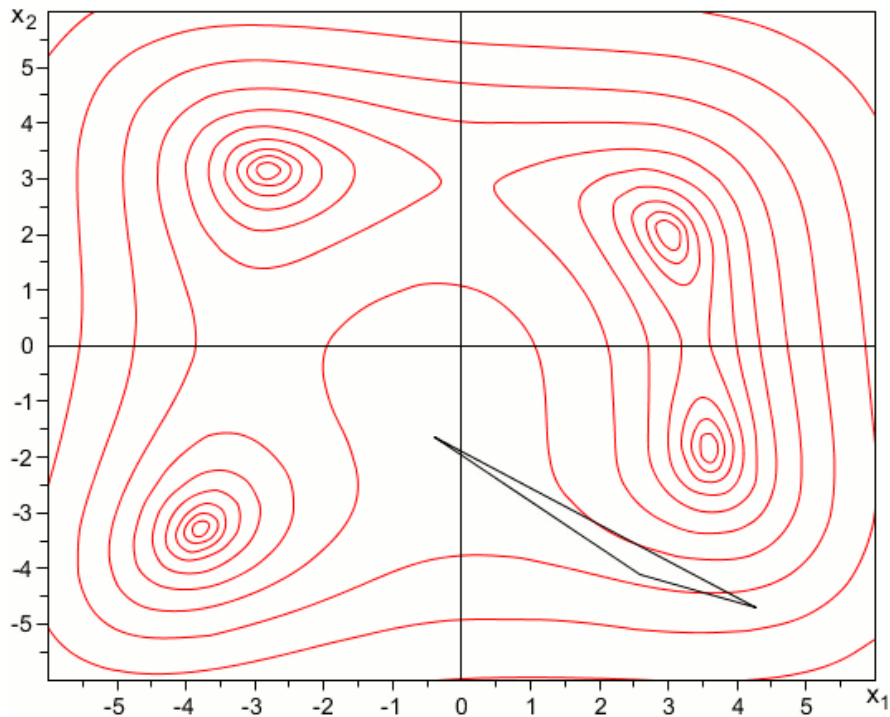
STEM offers a choice of Integration Engines

1. Finite Difference Solver (fast, good for demos)

2. Runge-Kutta-Feldberg (RKF45) *adaptive* integration
   - Adaptive step size
   - Very efficient
   - Synchronized across threads (multi-core engine)
   - Accurate
   - Computational error is estimated and controlled and an appropriate step size set automatically

3. DormandPrince54 (from *apache.math.commons*)

4. … User Contributed Solvers
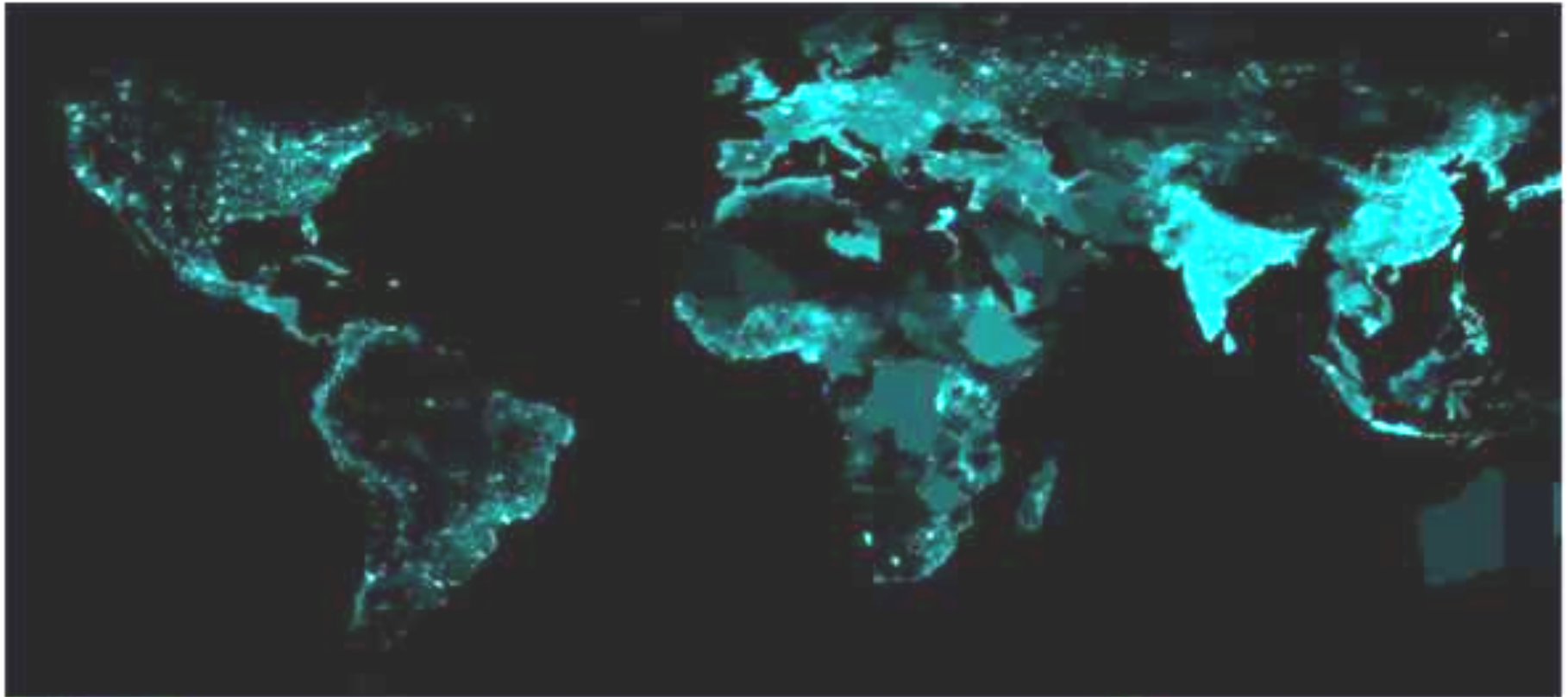
STEM supports global simulation

Global population data by admin region validated against *LandScan 2007™*

High Resolution Global Population Data Set
© UT-Battelle, LLC, operator of Oak Ridge National Laboratory
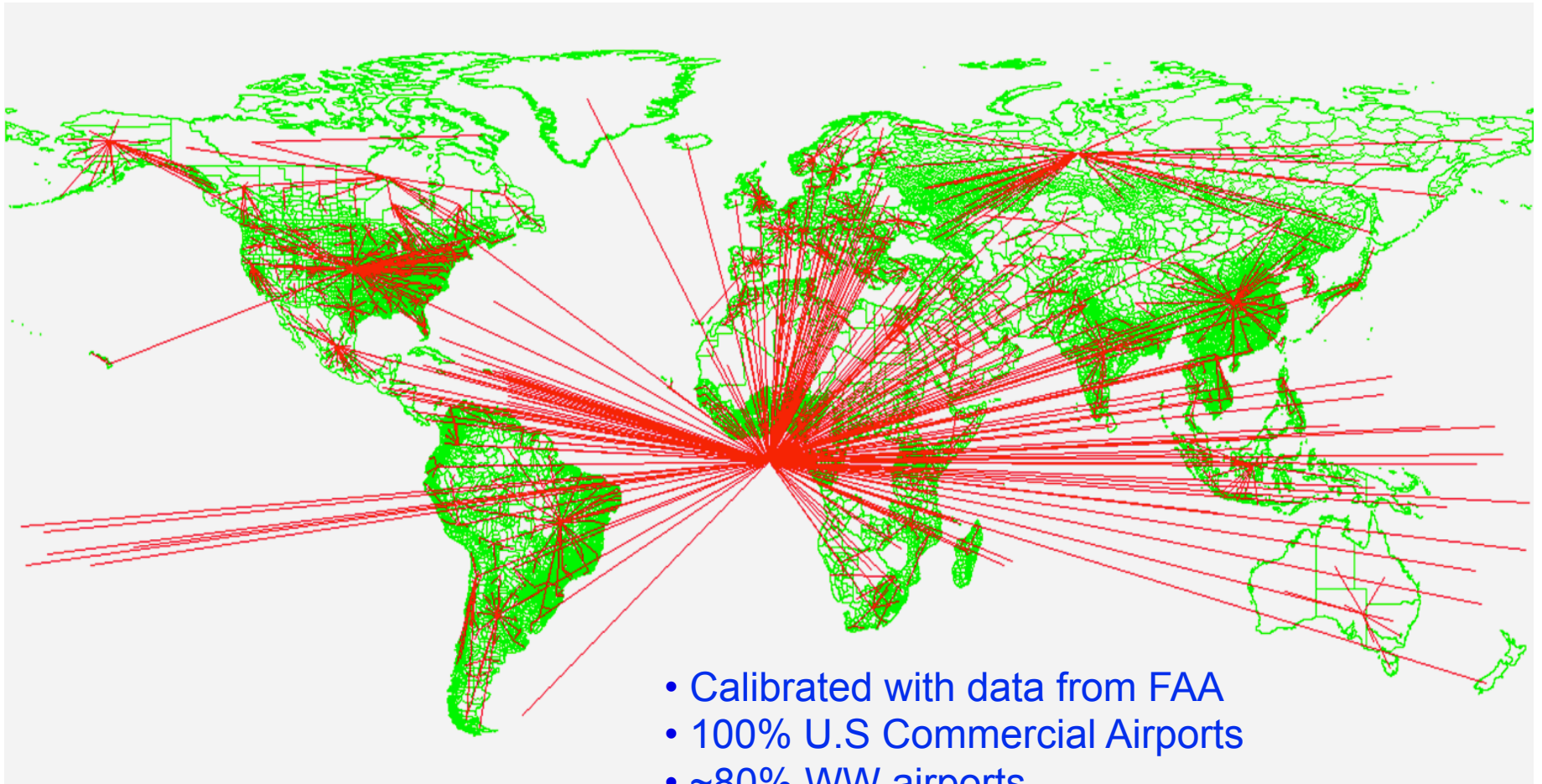
# Transportation Models
Global Air Travel

- Calibrated with data from FAA
- 100% U.S Commercial Airports
- ~80% WW airports
- Lessler J, Kaufman JH, Ford DA, Douglas JV. 2009. "The cost of simplifying air travel when modeling disease spread," PLoS ONE 4(2): e4403. doi: 10.1371

Other Features

- Ability to Run Lattice Graphs

- Ability to Import Custom Graphs (Pajek)

- Support for Zoonotic Disease (Multiple Populations)
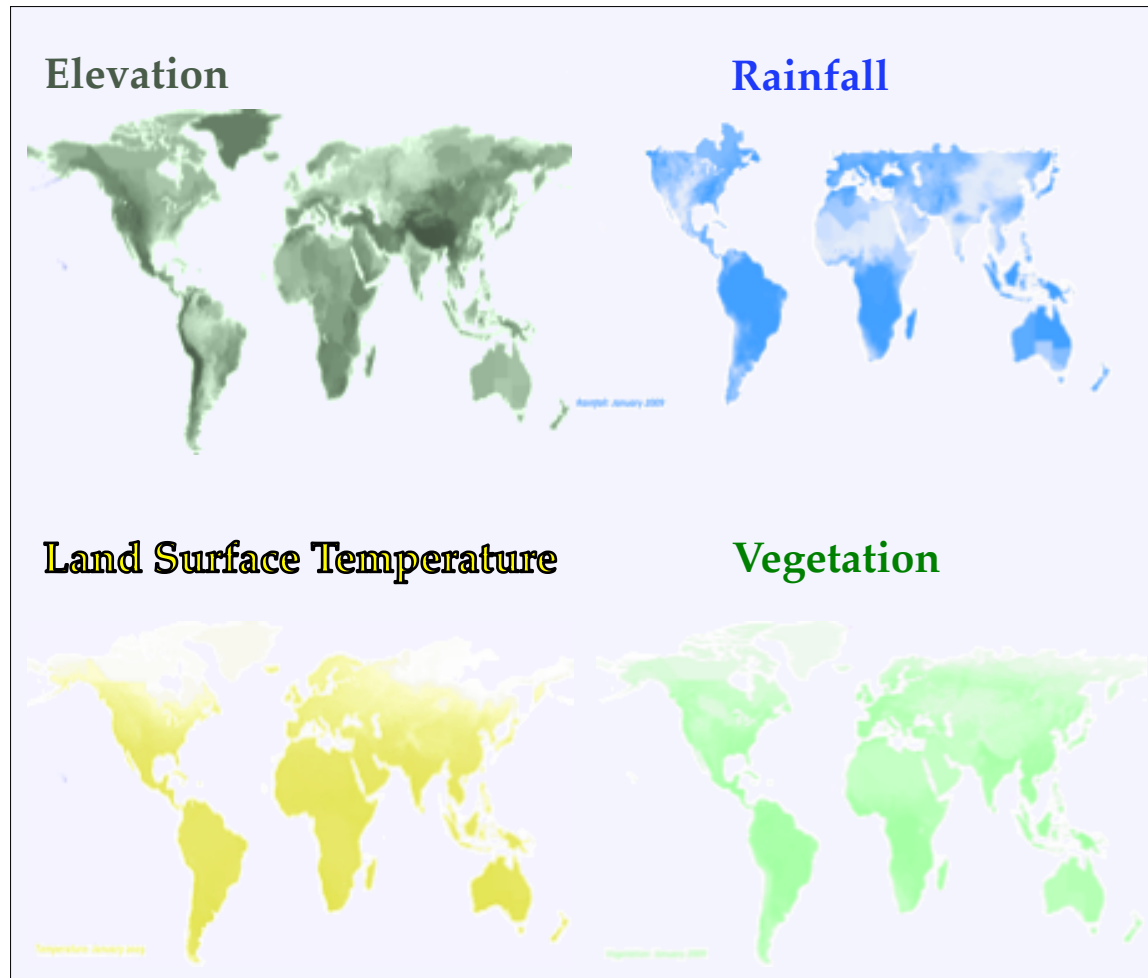   - Malaria
    - Yellow Fever
    - Dengue
    - etc.
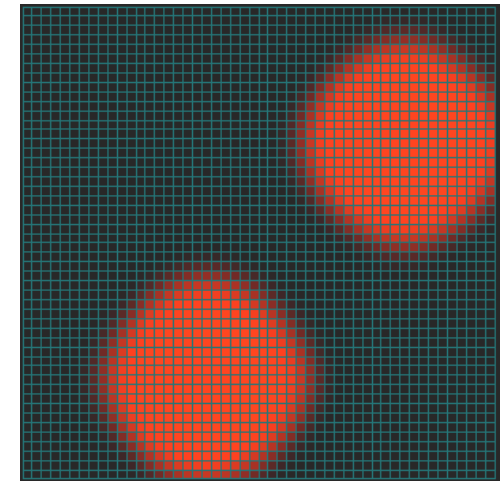
Global Earth Science Data
(courtesy of NOAA GLOBE, NASA Earth Observatory)

Elevation

Rainfall

Land Surface Temperature

Vegetation

Anopheles Mosquito Density
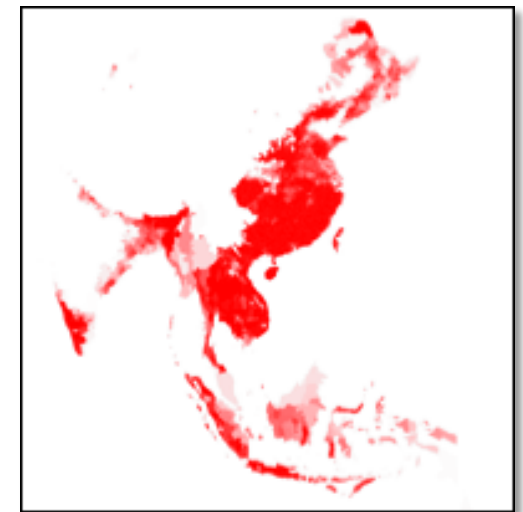
Comparing STEM Anopheles Probability Model to USGS VCAP model for Africa (2012)
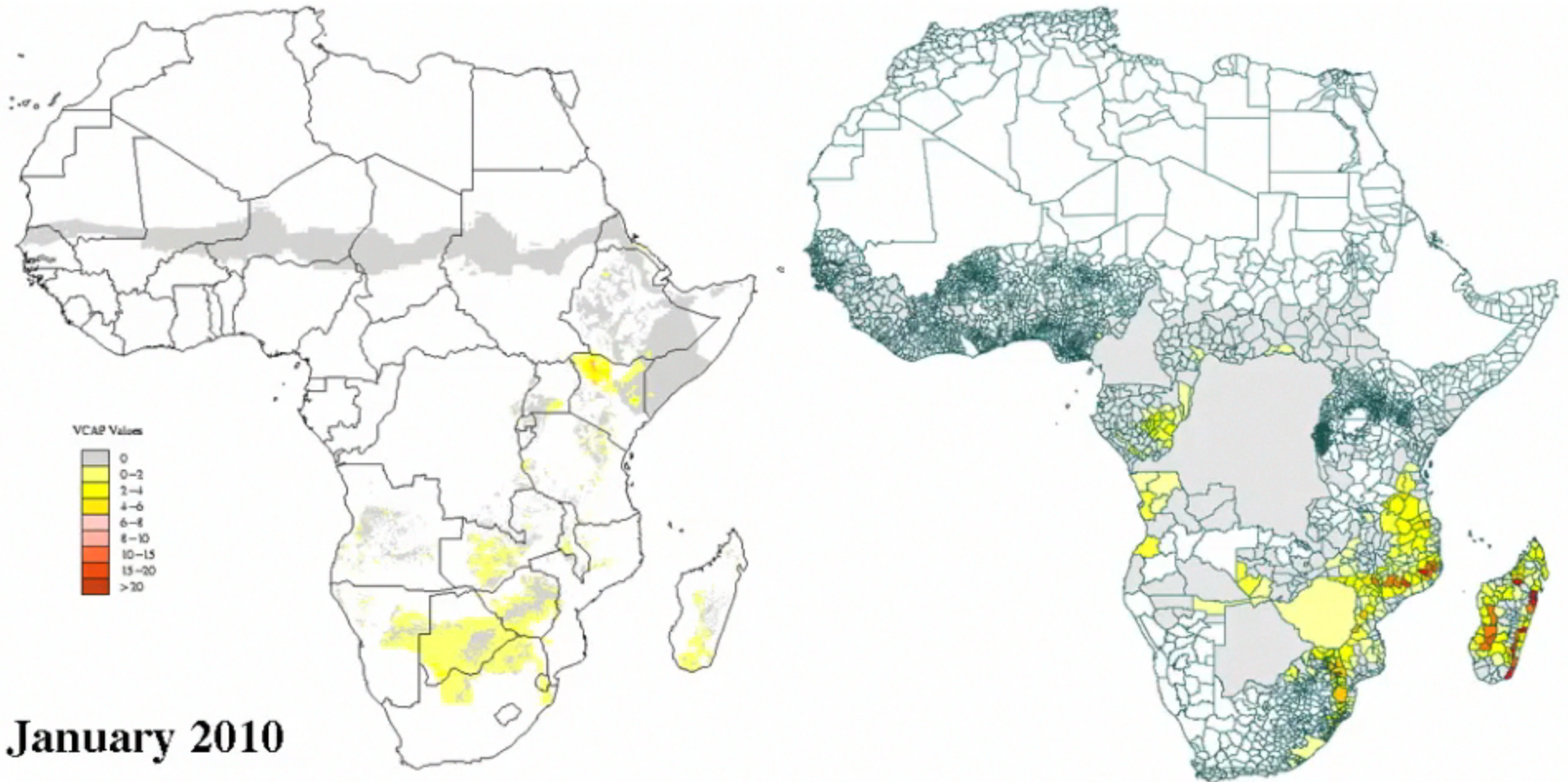
Model Validation and Analysis:

Data/Model

- **Correlating Time Series**
  - Comparing models to Models
  - Models to Data
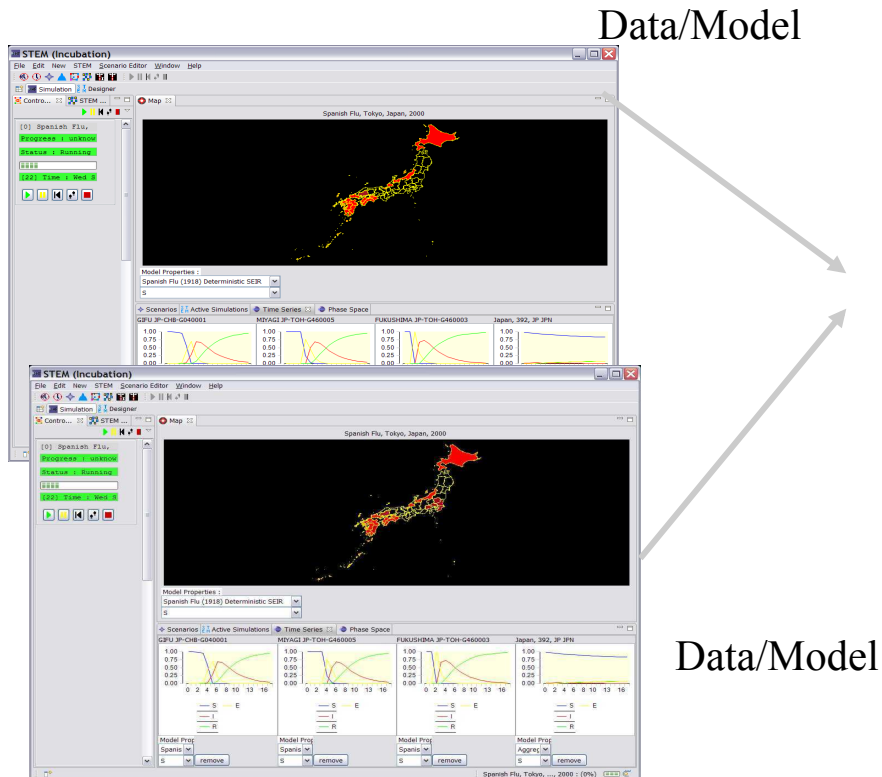
Error Function Analysis
Automated Experiments
Nedler-Mead Simplex
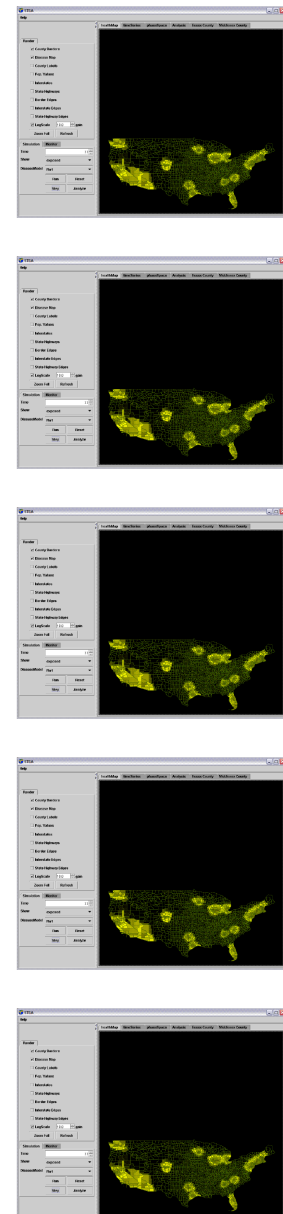Cross Validation

Data/Model

Analytic Measure

- Compare models to data and models to models
  - Quantify improvements in modeling
  - Quantify accuracy loss with time

Evaluation of Public Health Policies

HCN/HL7

Weather

Air Traffic
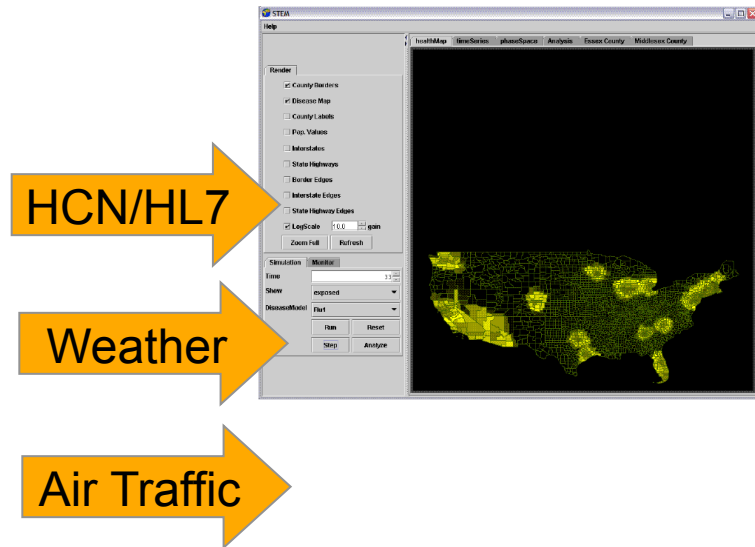
STEM includes a framework of
- •Predicates
- •Modifiers
- •Triggers

Masks

No Air travel

Vaccinate

Masks/No Air

?
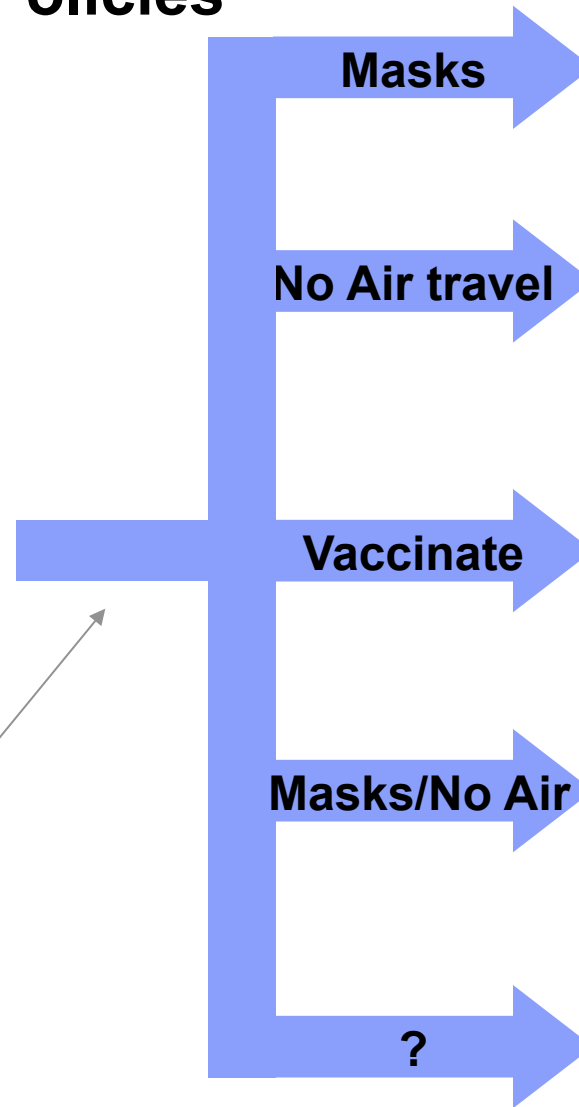
Modeling Influenza

J. Kaufman[1], S. Edlund[1], M. Bromberg[2], G. Chodick[3],
J. Lessler[4], J. Douglas[1], Z. Kaufman[2], A. Leventhal[6],
R. Marom[3], V. Shalev[3]

[1]*IBM Almaden Research Center, USA,*
[2]*Israel Center for Disease Control, Israel,*
[3]*Maccabi Health Care Services, Israel,*
[4]*Johns Hopkins School of Public Health, USA,*
[5]*IBM Haifa Research Center, Israel;*
[6]*Israel Ministry of Health, Israel*

August 2012

Simulations Based on ICDC Influenza Data
as initial condition
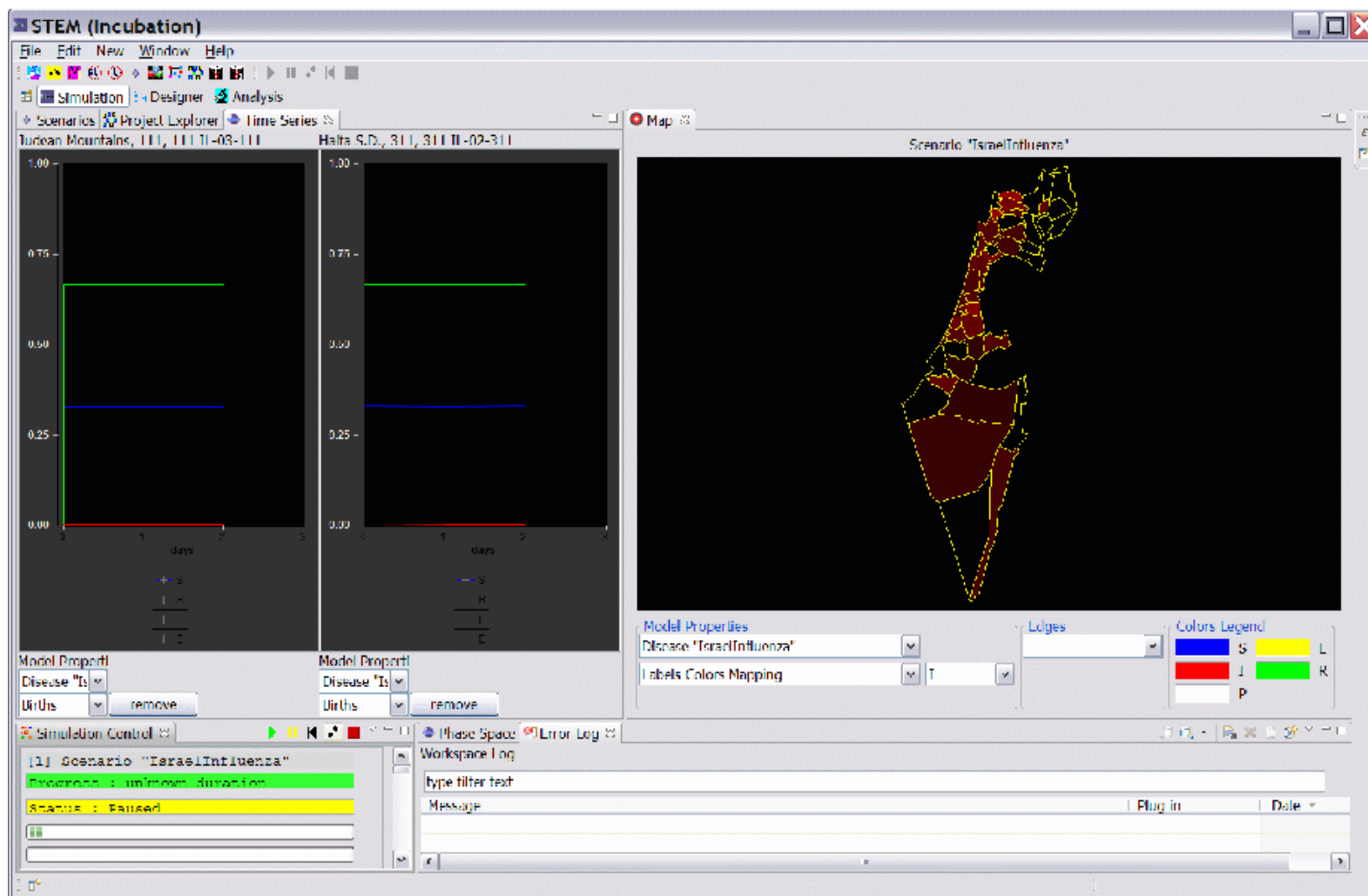
The Data

- **Data provided by the Israel Center for Disease Control (ICDC)**

  – Originated from Maccabi Health Care Services, 2nd largest HMO in Israel serving approx. 25 % of population

  – 10 years of summarized daily case reports of "Influenza Like Illness" (ILI)

  – Mapped to 49 administrative regions of the 15 Israeli sub-districts

Testing Three Hypotheses about Influenza Transmission

Using this new model, test three hypotheses in terms of β_i, ρ_i and σ_i.

- Hypothesis 0: Transmission identical in both A and B dominant years
 ($\beta_i = \beta_o$, $\sigma_i = \sigma_o$, $\rho_i = \rho_o$ for all years). 7 parameters

- Hypothesis 1: Maximum rate of transmission varies between A and B dominant years.
  $\beta_{i,,} = \beta^A$ in A dominant seasons, and $\beta_{i,,} = \beta^B$ in B dominant seasons
  ($\sigma_i = \sigma_o$, $\rho_i = \rho_o$ for all seasons). 8 parameters

- Hypothesis 2: The maximum rate of transmission and the effect of seasonal forcing vary between influenza A and influenza B. 10 parameters
  $\beta_{i,,} = \beta^A$, $\sigma_{i,,} = \sigma^A$, $\rho_{i,,} = \rho^A$ in A dominant years
  $\beta_{i,,} = \beta^B$, $\sigma_{i,,} = \sigma^B$, $\rho_{i,,} = \rho^B$ in B dominant years.

Cross Validation as a Measure of Predictive Ability
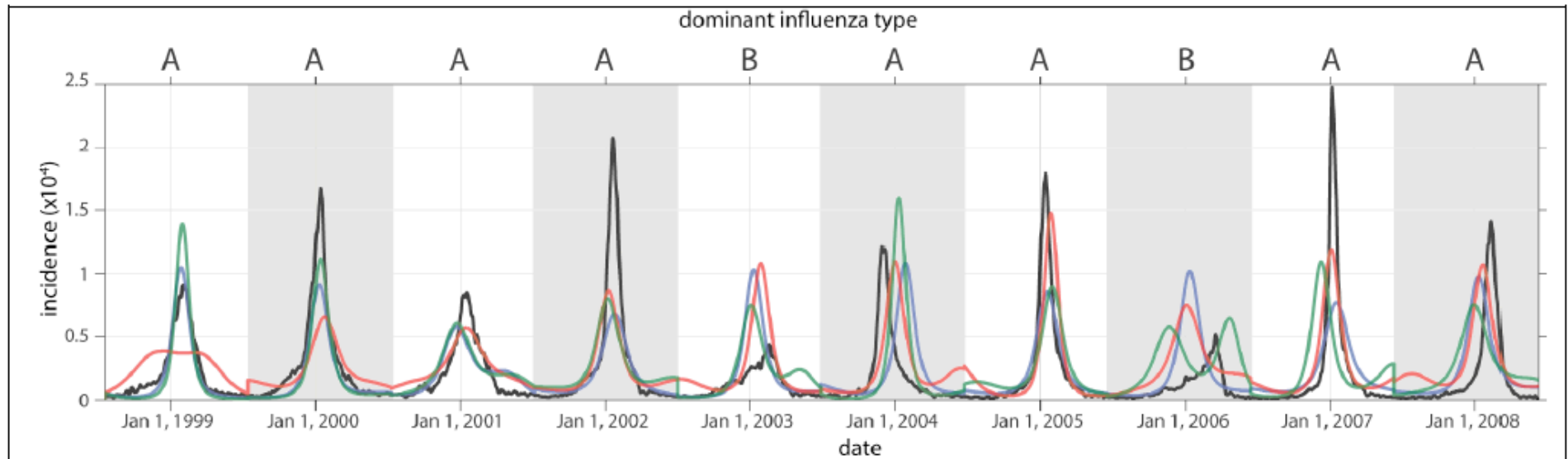
Parameter	lowest	typical	upper	H0	H1(A,B)	H2(A,B)
Period of Infection [days]	1.0	2	4.8	1.1	1.1	1.1
Period of Immunity [years]	1	4-6	8	5.9	4.9	5.4
Reproductive Number	1.1	1.5-2.0	20	2.6	2.8,2.7	2.6,2.6
NRMSE	-	-	-	0.244	0.238	0.249

H1N1: *Work With the the Mexico Ministry of Health* Gobierno del Distrito Federal (GDF)

IBM Research working with the IBM Foundation, the office of Corporate Responsibility, and IBM GBS, delivered new Servers with DB2, Websphere, PHIAD and STEM installed to the Mexican Ministry of I (GDF). Tutorials and workshops were held in Mexico for the GDF Dir General de Planeación y Coordinación and his colleagues and to Instituto Mexicano del Seguro Social (IMSS).

"…it is extremely pleasant to meet a Company that offers their help in these difficult times with a solution that perfectly fits our needs and that does not expect to take advantage of the situation we are living"

Lic Luis Guillot Duenas   Director Ejecutivo de Evaluación y Seguridad de las Tecnologías de la Información - Contraloria General
Lic Josune Arceluz de Diego   Directora de Gestion Gubernamental para la Atención Ciudadana
Lic Irak Lopez Davila   Coordinador General de Modernización Adminitrativa
Lic. Alejandra Olguin
Lic Luis Alfonso Caso   Director General de Planeación y Coordinación Sectorial

August 2012

Recent H1N1 Outbreak
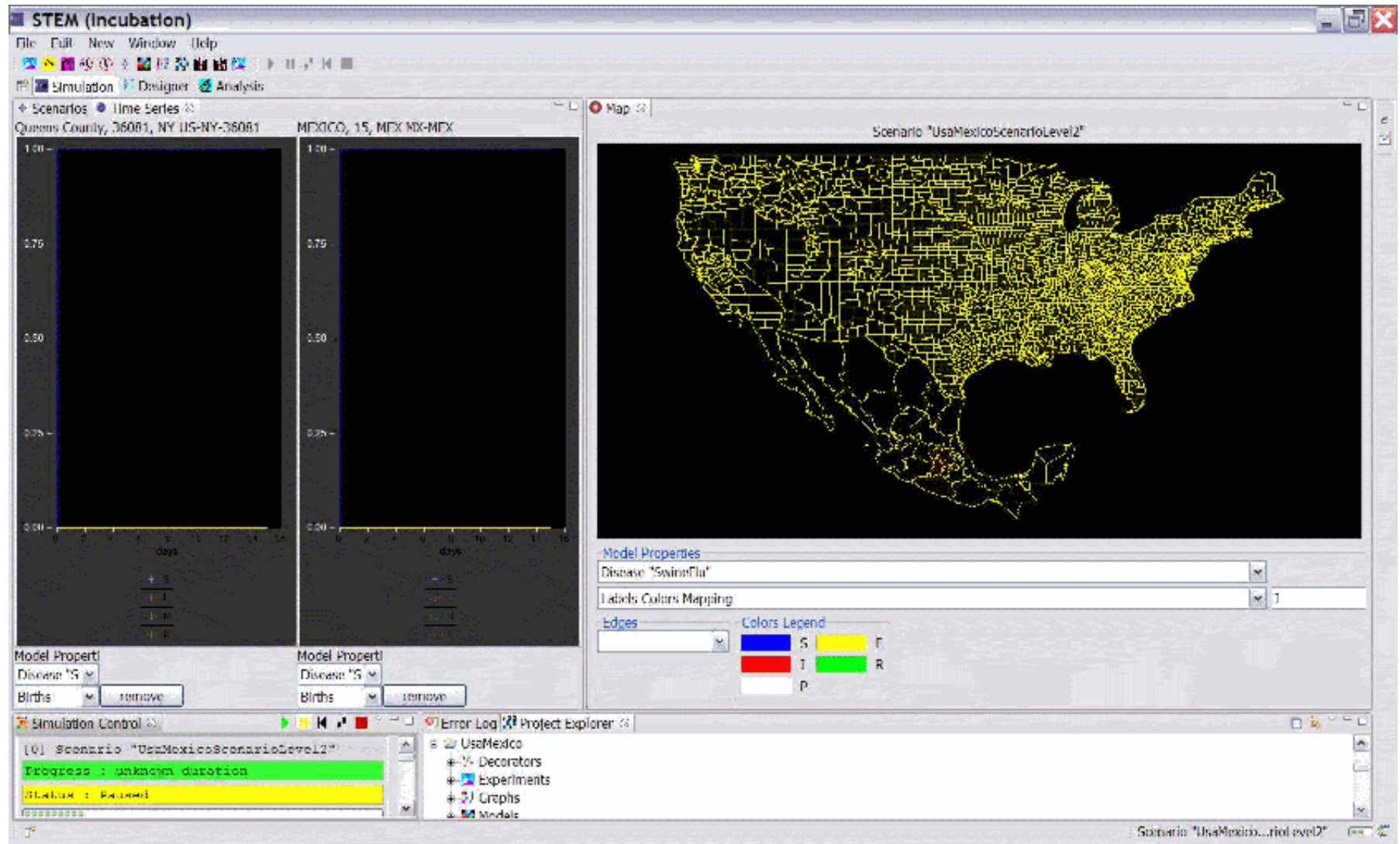
H1N1

- What is the reproduction number Ro ?

- Case report data not complete

- What could we learn from public data
  - When first cases (~10) showed up in NY there were thousands in Mexico city
  - Flu in Mexico peaked shortly after spread to air transport system

- Can we use the spread by air transport as an indicator for Ro ?

First Experiment (very little data available at the time)

- Run Simulation as a function of Ro
- Use STEM air travel plug-in to model connection between Mexico (GDF) and USA
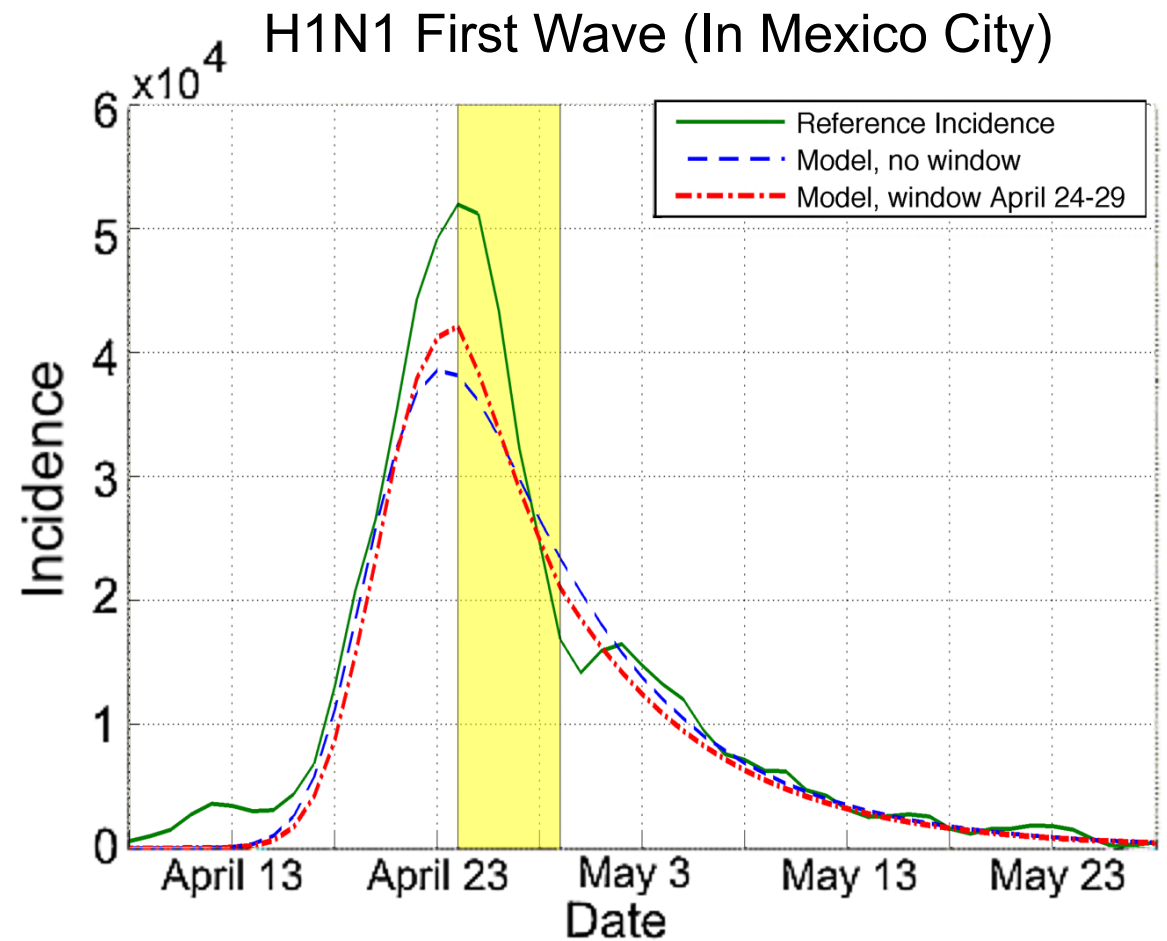- Plot Incidence in Mexico City at a function of Ro when I=10 in NYC
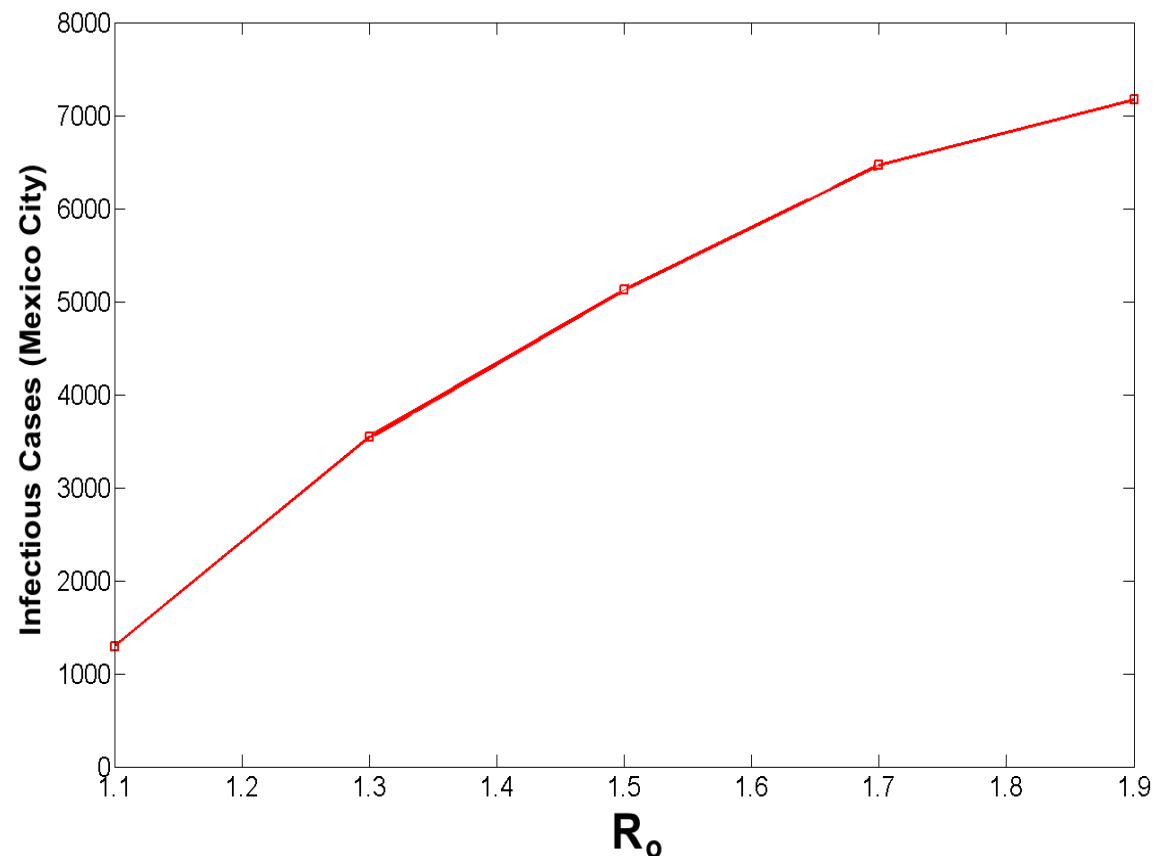
Infectious Count in Mexico City When I(NY)~10

Hypothesis Testing
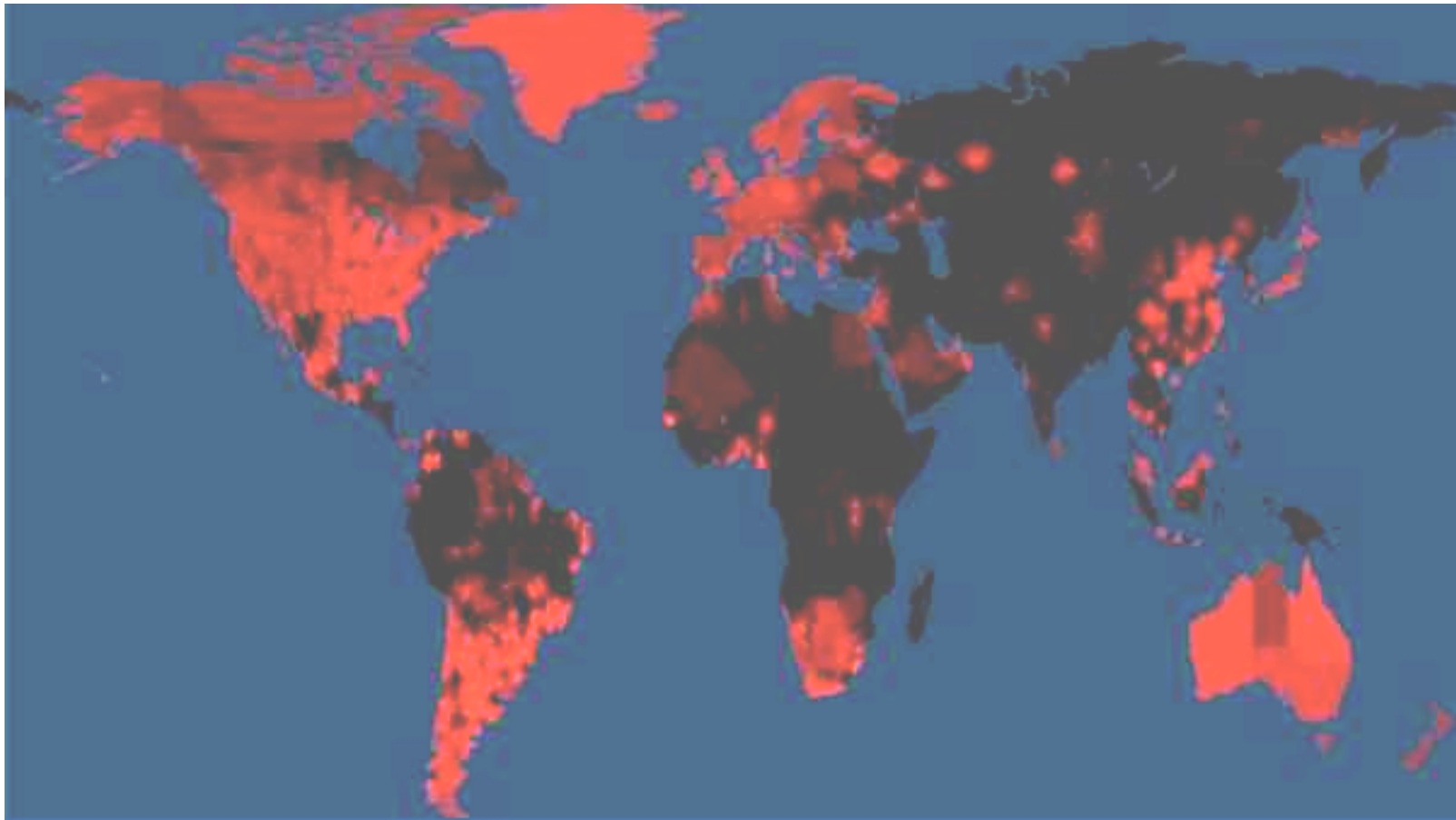
Nelder-Mead Experiment
"discovers" 5 day window
where schools were closed.

Transmission was reduced by
22%

H1N1 First Wave (In Mexico City)

Global scale influenza simulation (STEM on server)

Vector Disease Models

Malaria
MacDonald-Ross model extended by Aaron and May

$$\frac{ds_{\vec{r}}(t)}{dt} = -ab\frac{\hat{N}(t,\vec{r})}{N(\vec{r})}\hat{i}(t,\vec{r})s(t,\vec{r}) + \alpha r(t,\vec{r})$$

$$\frac{de_{\vec{r}}(t)}{dt} = ab\frac{\hat{N}(t,\vec{r})}{N(\vec{r})}\hat{i}(t,\vec{r})s(t,\vec{r}) - \varepsilon e(t,\vec{r})$$

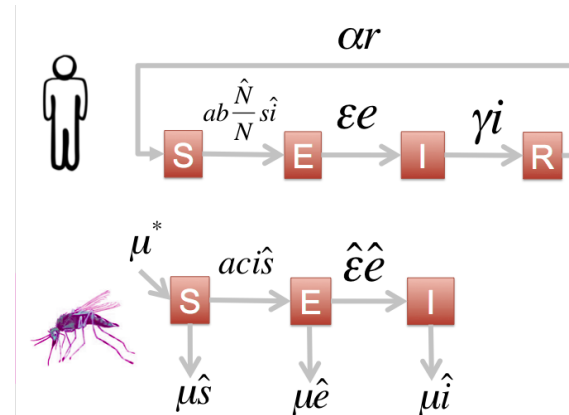$$\frac{di_{\vec{r}}(t)}{dt} = \varepsilon e(t,\vec{r}) - \gamma i(t,\vec{r})$$

$$\frac{dr_{\vec{r}}(t)}{dt} = \gamma i(t,\vec{r}) - \alpha r(t,\vec{r})$$

humans

$$\frac{d\hat{s}_{\vec{r}}(t)}{dt} = -aci(t,\vec{r})\hat{s}(t,\vec{r}) + \mu^* - \mu\hat{s}(t,\vec{r})$$

$$\frac{d\hat{e}_{\vec{r}}(t)}{dt} = aci(t,\vec{r})\hat{s}(t,\vec{r}) - \hat{\varepsilon}\hat{e}_{\vec{r}}(t) - \mu\hat{e}(t,\vec{r})$$

$$\frac{d\hat{i}_{\vec{r}}(t)}{dt} = \hat{\varepsilon}\hat{e}_{\vec{r}}(t) - \mu\hat{i}(t,\vec{r})$$

Anopheles

Parameter	Description	Parameter Value
	Size of female mosquito population	Calculated[1]
N	Size of human population	From census data
a	Biting rate (bites/day)	0.008[1]
b	Prop. inf. bites on human causing inf.	1.0
γ	Human recovery rate	0.00725
c	Prop. susc. bites on human causing inf.	1.0
μ^*	Mosquito birth rate	0.071[2] (1/days)
μ	Mosquito death rate	0.071[2] (1/days)
$1/\varepsilon$	Latent period human	12 days
α	Immunity loss rate (human)	0.00207 (483 days)
$1/\hat{\varepsilon}$	Latent period mosquito	111/(T-16) (*P. Falc.*)

# The Importance of *Fluctuations*
rather than absolute incidence

•Reporting of malaria incidence has a high uncertainty.
•Reporting rates vary by country.
•WHO estimates and Malaria Atlas differ by a factor of 4 !!

•**Question 1**: During years when *global* malaria burden is higher than average, which regions contribute most to the increase?
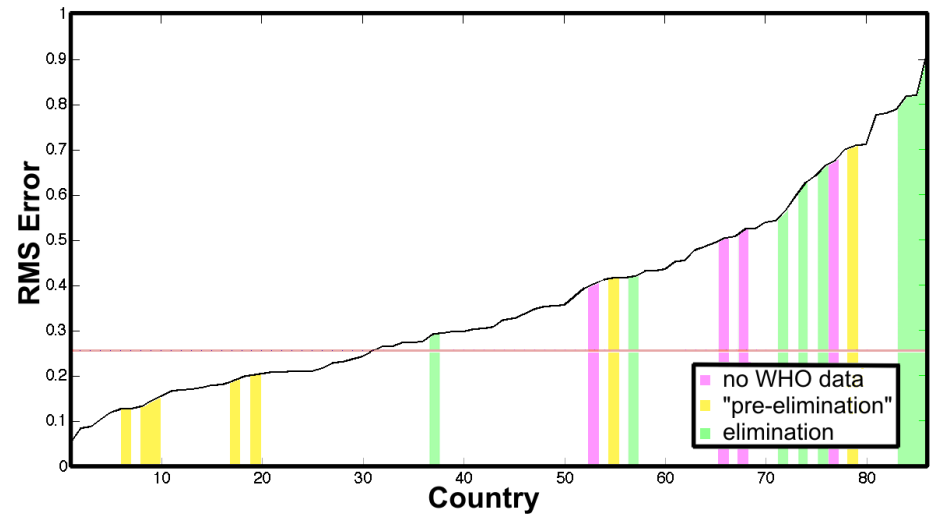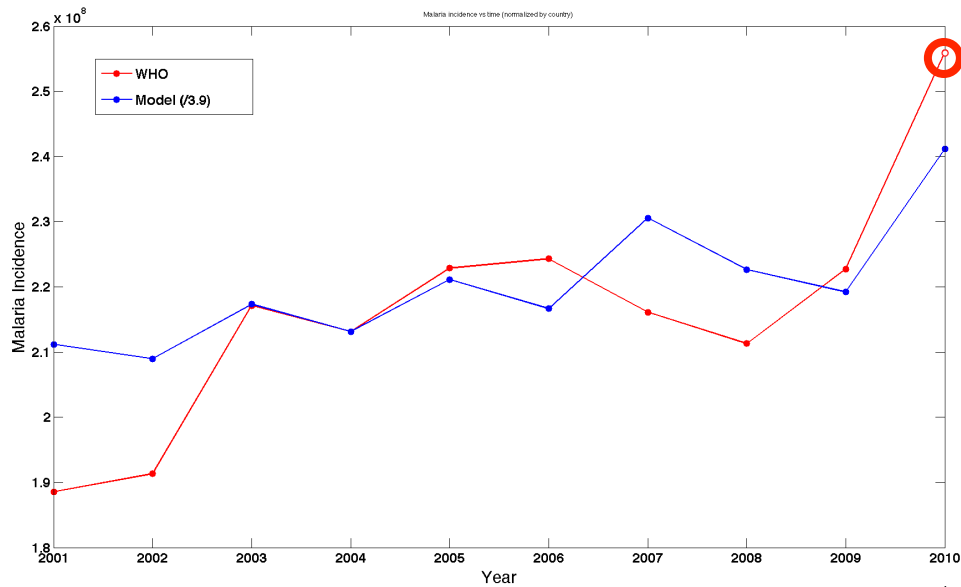
• **Question 2**: How does malaria burden fluctuate with *local* variations in temperature
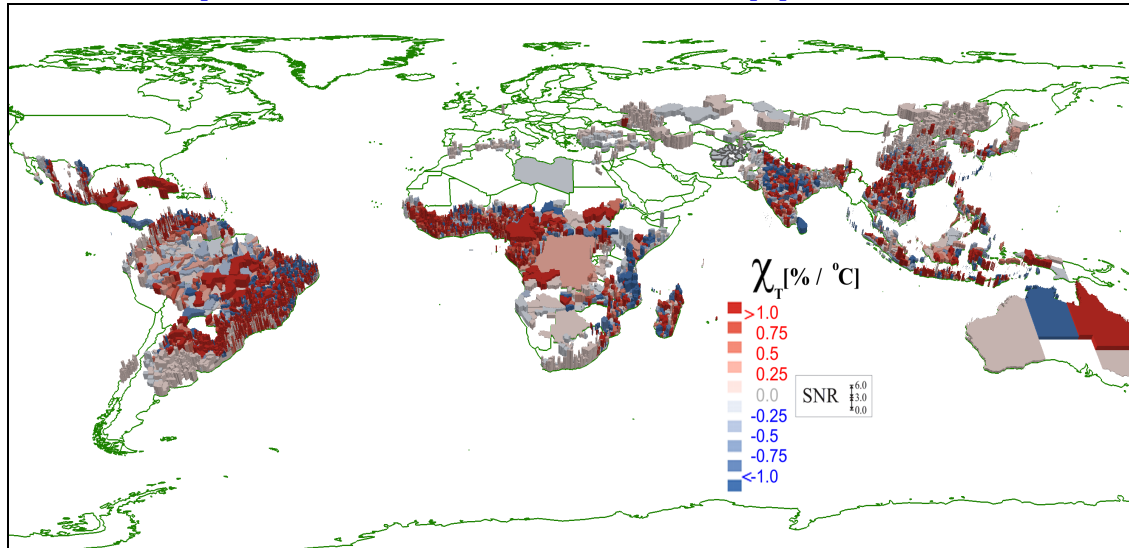and precipitation?

See animation….

$$\chi_I(\vec{r}) = \frac{1}{n_a} \sum_a \sum_b \frac{I_a(\vec{r}) - I_b(\vec{r})}{N(\vec{r})n_b}$$
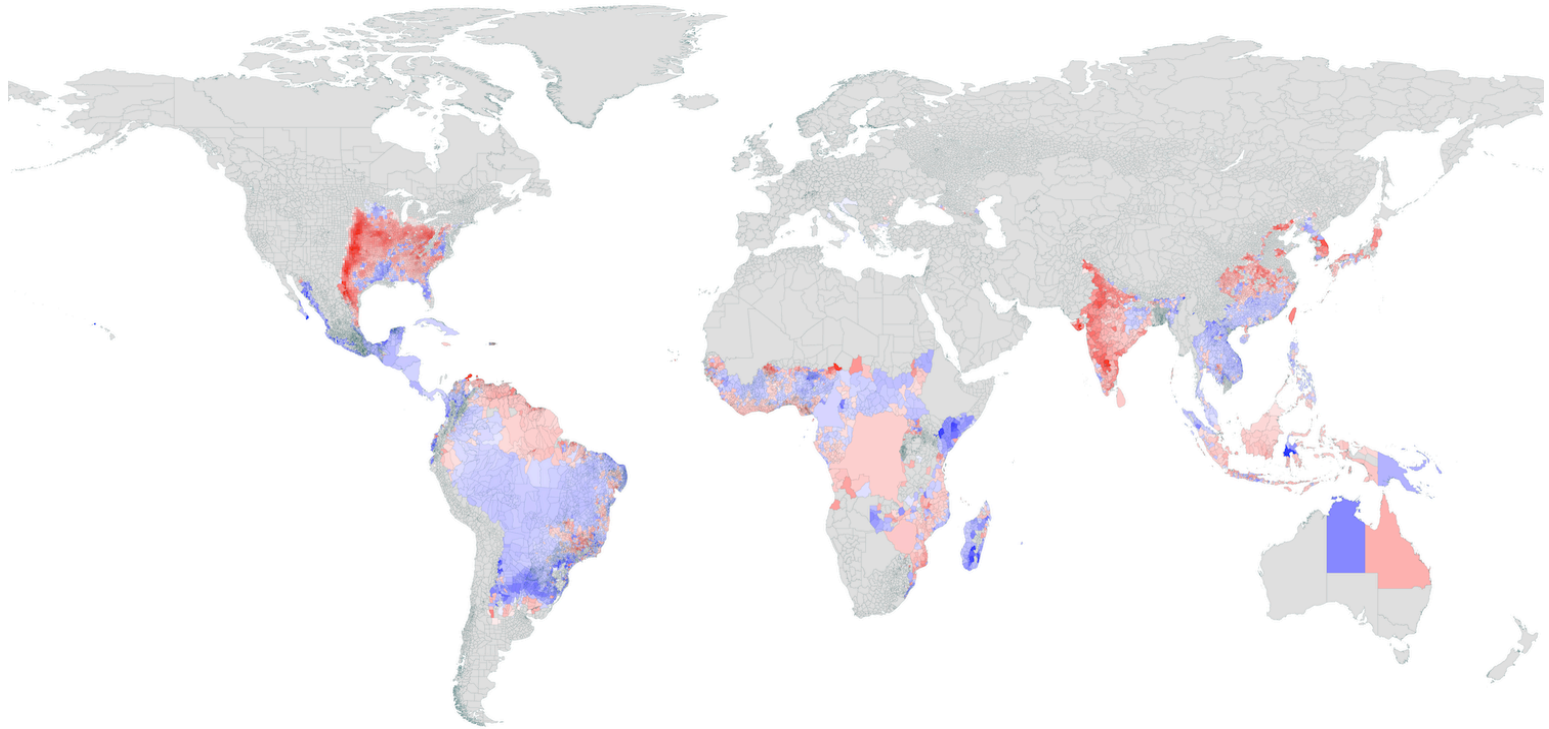
Predicting changes in global incidence

Malaria Response to Precipitation and Temperature
response function can be applied to model and/or to W.H.O. data !!

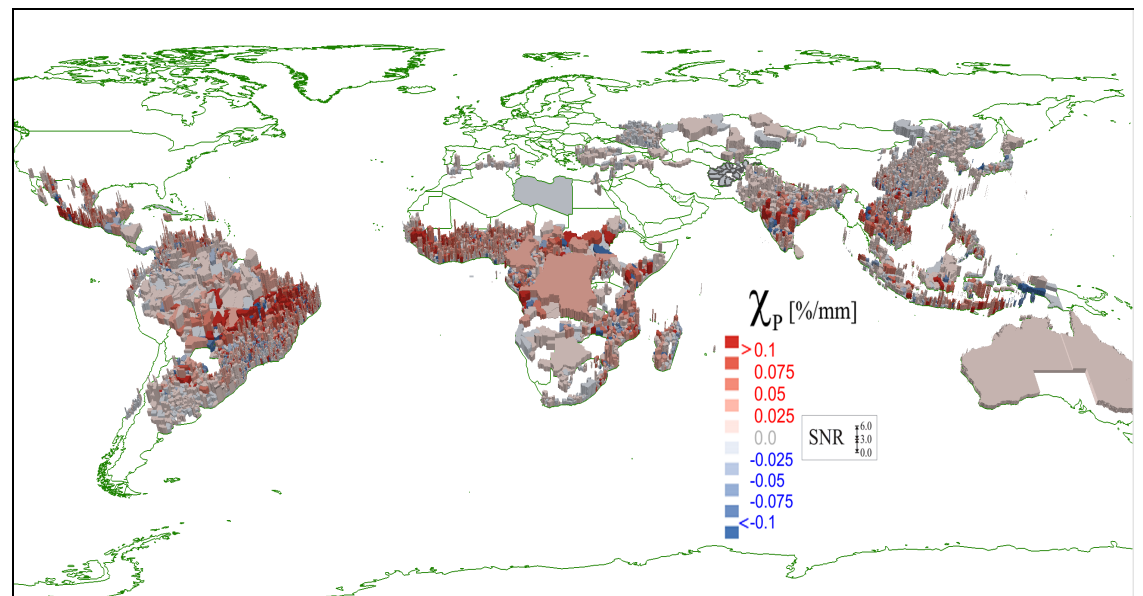$$\sigma_T(\vec{r},a,b) = \frac{I_a(\vec{r}) - I_b(\vec{r})}{[T_a(\vec{r}) - T_b(\vec{r})]N(\vec{r})}$$

$$\sigma_P(\vec{r},a,b) = \frac{I_a(\vec{r}) - I_b(\vec{r})}{[P_a(\vec{r}) - P_b(\vec{r})]N(\vec{r})}$$

$$S_T(\vec{r}) = \left\{ \sigma_T(\vec{r},a,b) \middle| \forall a \in C, \forall b \in C, a > b \right\}$$

$$S_P(\vec{r}) = \left\{ \sigma_P(\vec{r},a,b) \middle| \forall a \in C, \forall b \in C, a > b \right\}$$

$$\chi_T(\vec{r}) = \overline{S_T(\vec{r})}$$

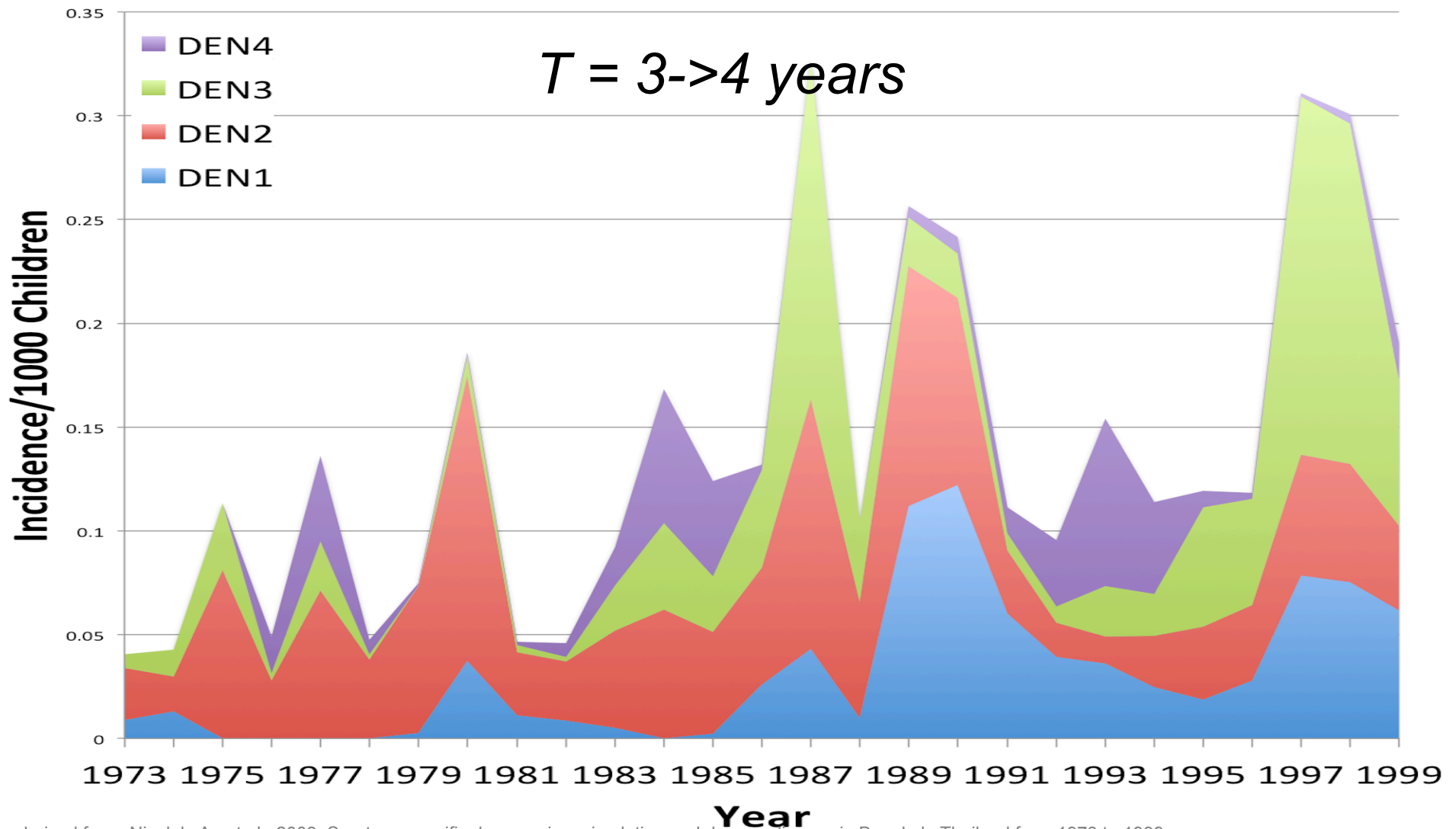$$\chi_P(\vec{r}) = \overline{S_P(\vec{r})}$$

Dengue Fever and Dengue Hemorrhagic Fever

- Dengue is a mosquito-borne infection (vector disease)

- 2.5 billion people at risk

- A leading cause of death among *children* in Asia.

Quasi-periodic Dengue Outbreaks in Thailand

T = 3->4 years

derived from: Nisalak, A., et al., 2003. Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. American Journal of Tropical Medicine and Hygiene 68, 191-202.
Population growth rate ~ 2.6%/year

Antibody Dependent Enhancement (ADE)

- Antibodies are *normally* a good thing.

- ADE => non-neutralizing antiviral antibodies lead to
  - enhanced viremia
  - enhanced entry into host cells
  - increased infectivity
  - in the host cells.

- Increases risk of
  - dengue shock syndrom
  - dengue hemorrhagic fever
  - Death

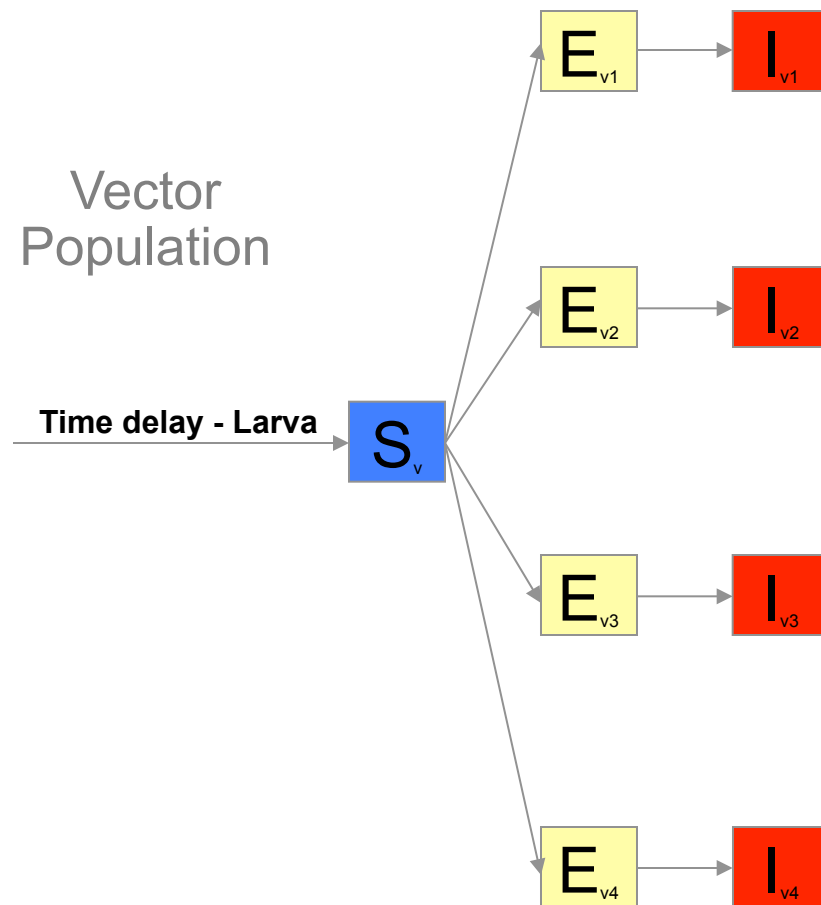- Has now reached Florida, Texas

- Endemic in Australia
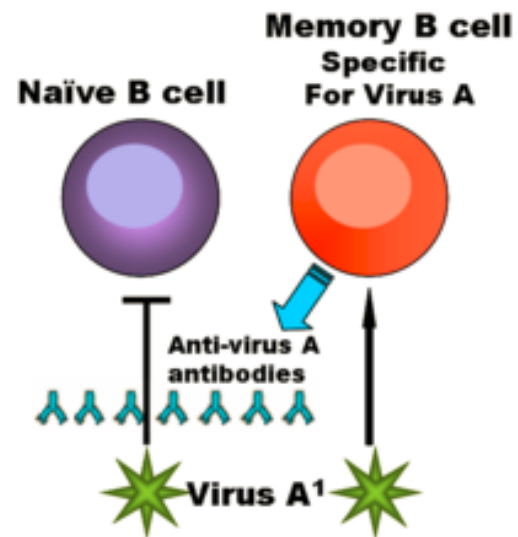
Comparing Three models of Dengue Fever

- **Model 1: Literature:** *Bianco et al.*

  - Leaves out the vector

  - Ignores host incubation

- **Model 2: Add the *Aedes* vector**

- **Model 3: Add incubation (the "full model")**

Full Dengue Host/Vector Model require 51 Compartments
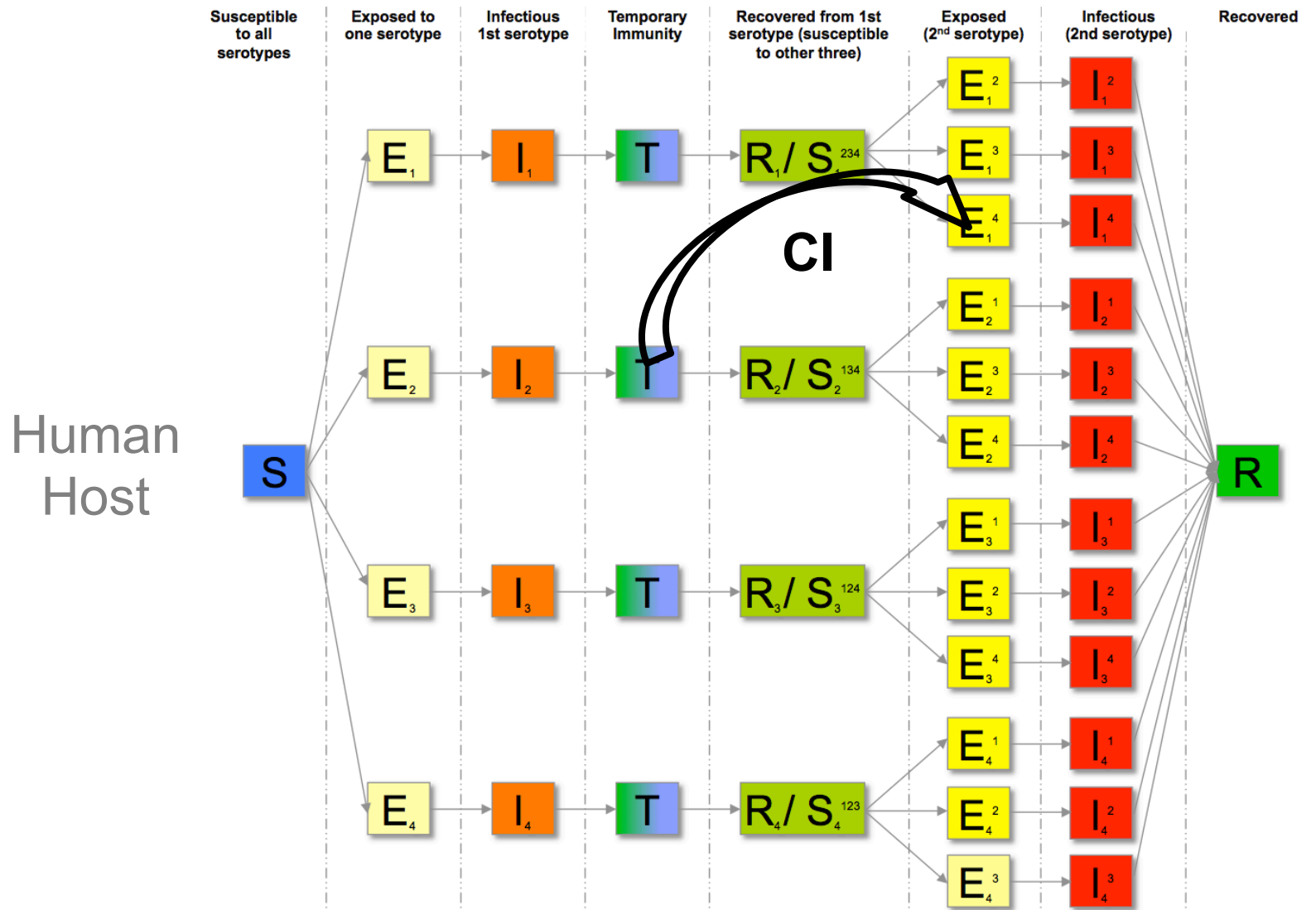
Vector Population Compartment Model

Epidemiological Parameters

Model input / Parameter	Value	Range	Sources (ref.)	Unit
R_0, basic reproductive number	2	2-4	(Billings et al., 2007; Chowell et al., 2007; Koopman et al., 1991)	dimension-less
μ^h, host birth and death rate	0.02	0.03-0.01	(Ferguson et al., 1999a)	yr^{-1}
μ^v, vector birth and death rate	35	24-61	(Chowell et al., 2007; Muir and Kay, 1998)	yr^{-1}
γ^h, host incubation rate	65	52-91	(Chowell et al., 2007)	yr^{-1}
γ^v, vector incubation rate	30	23-33	(Chowell et al., 2007)	yr^{-1}
β^h, transmission rate from host to host (no vector in model)	200	100-300	(Ferguson et al., 1999b)	yr^{-1}
β^{vh}, transmission rate (from vector) to host[*]	15	12-21	(Chowell et al., 2007)	yr^{-1}
β^{hv}, transmission rate (from host) to vector[*]	530	136-1000	(Chowell et al., 2007)	yr^{-1}
σ^h, host recovery rate	100	50-200	(Ferguson et al., 1999a; Gubler et al., 1981)	yr^{-1}
θ^h, cross immunity loss rate of host population	2	2-9	(Wearing and Rohani, 2006)	yr^{-1}
m, vector per host	1.9[#]	1-6	(Chowell et al., 2007, p.13)	dimension-less
ϕ, ADE factor	n.a.	1-3	Parameter range explored	dimension-less
ε, strength of cross immunity	n.a.	0-1	Parameter range explored	dimension-less

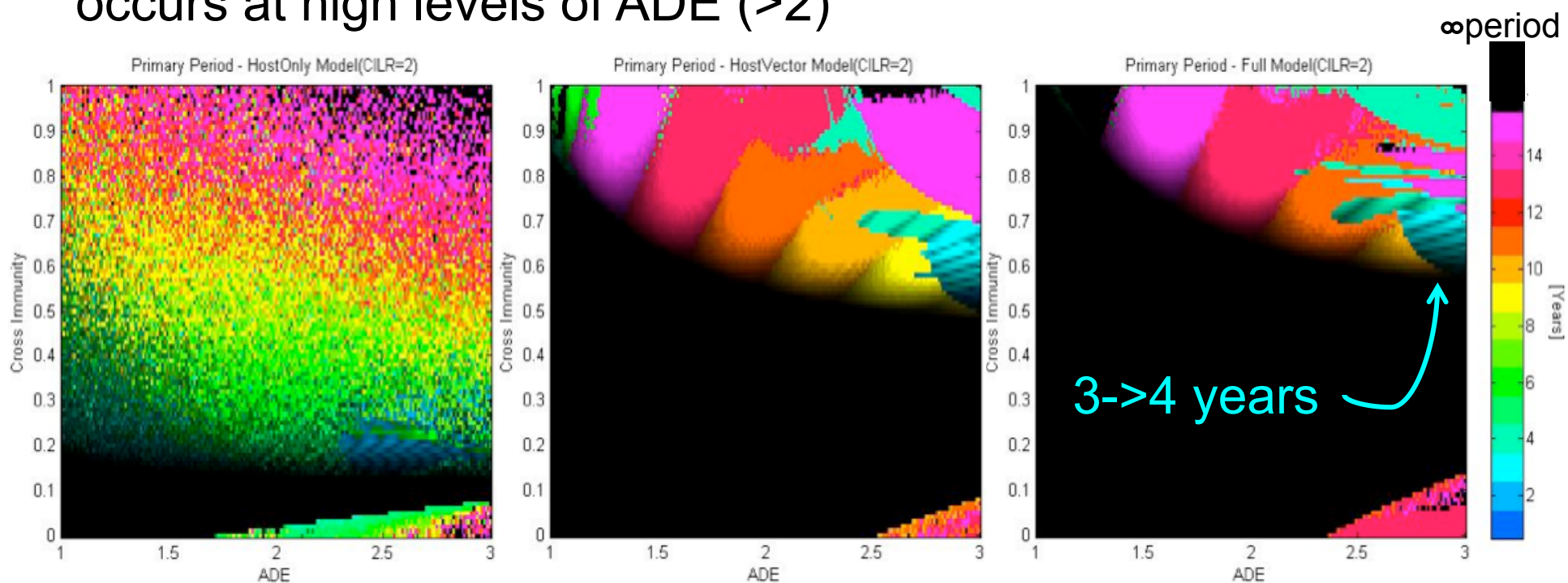# Onset of Dynamic Behavior (the Hopf Bifurcation)
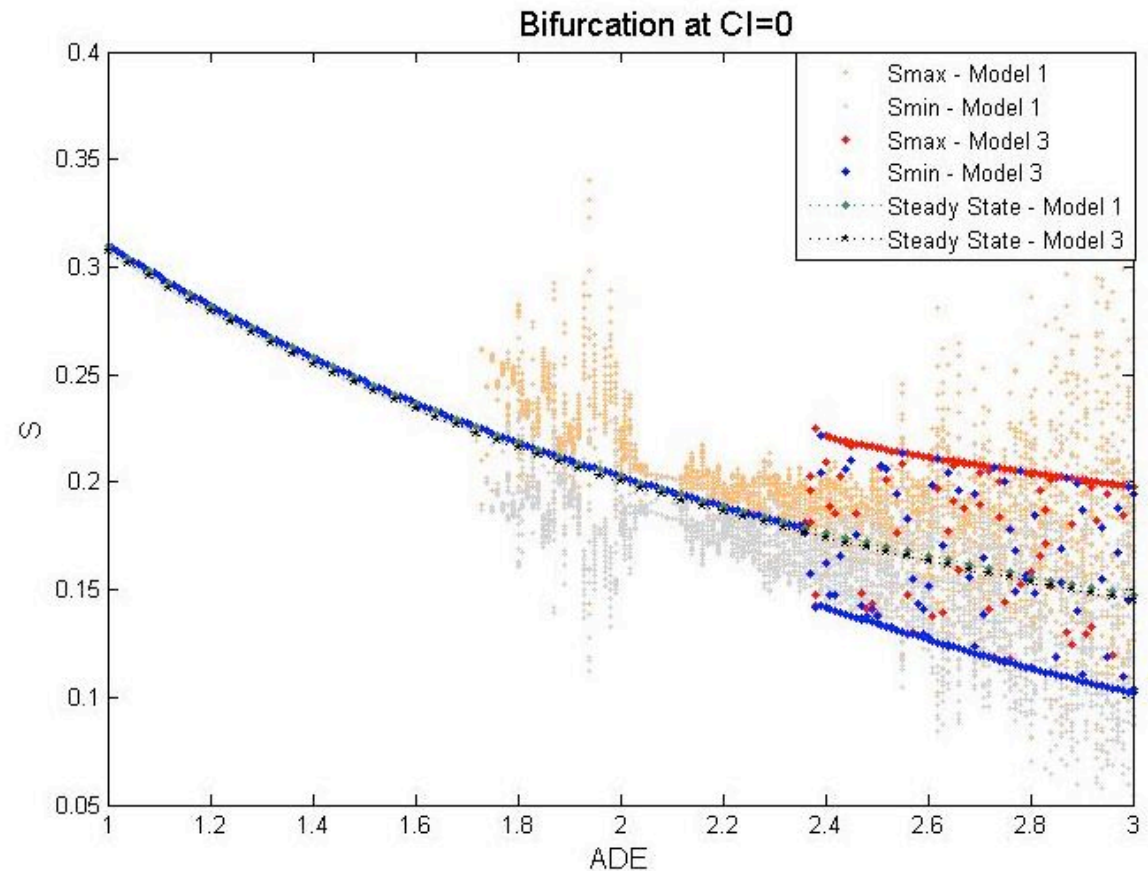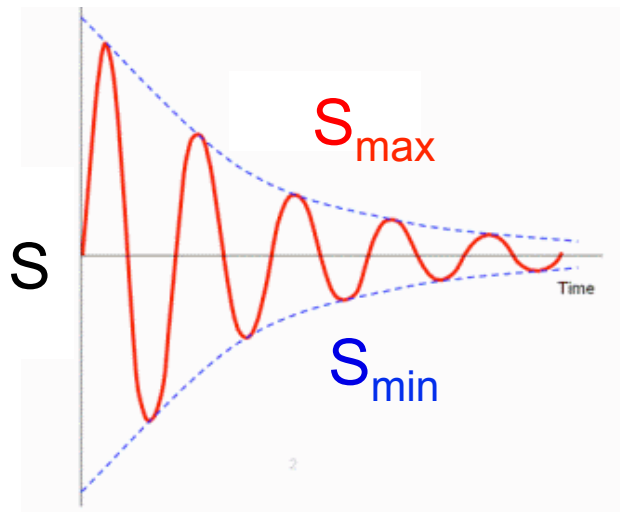
- Analyzing the power spectrum of S(t) we can simultaneously
  - Detect the hopf bifurcation
  - Measure dominant frequency (or period)

ADE vs Cross Immunity:
Explaining the observed 3-4 year quasi-periodic outbreaks in Thailand

A realistic period for dengue outbreaks (3-4 years)
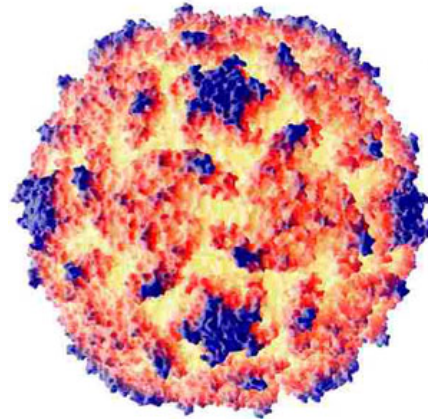occurs at high levels of ADE (>2)

Each image (map of phase space) required 10,000 simulations of 9000 years each
Paper in preparation for J. Theor Biology

Polio

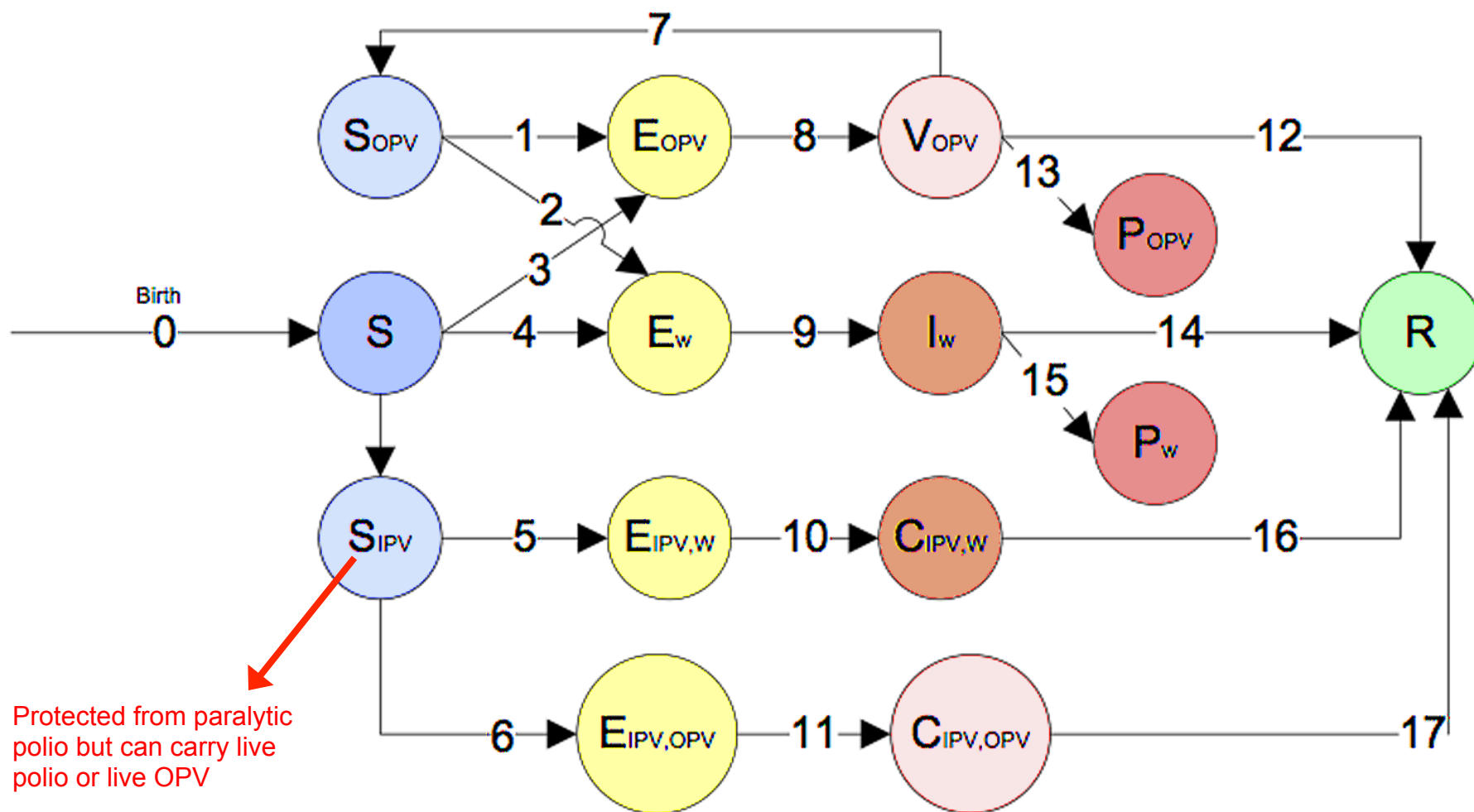**Coordinating Oral Polio Vaccine (OPV) and Intravenous Polio Vaccine (IPV) strategies**

http://wiki.eclipse.org/Polio_Disease_Transmission_Model
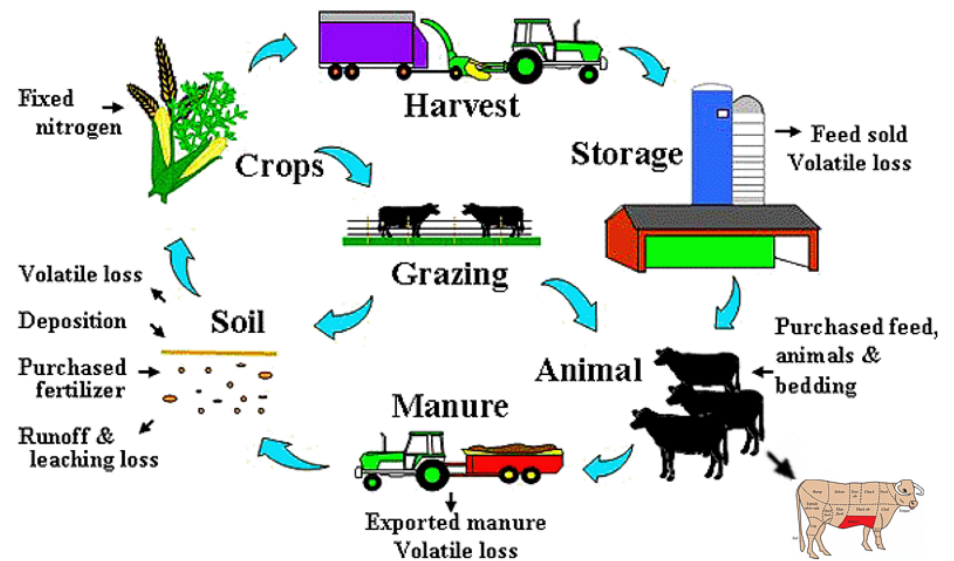
planned for STEM 1.4.0

Polio vaccination: OPV vs IPV

Protected from paralytic polio but can carry live polio or live OPV

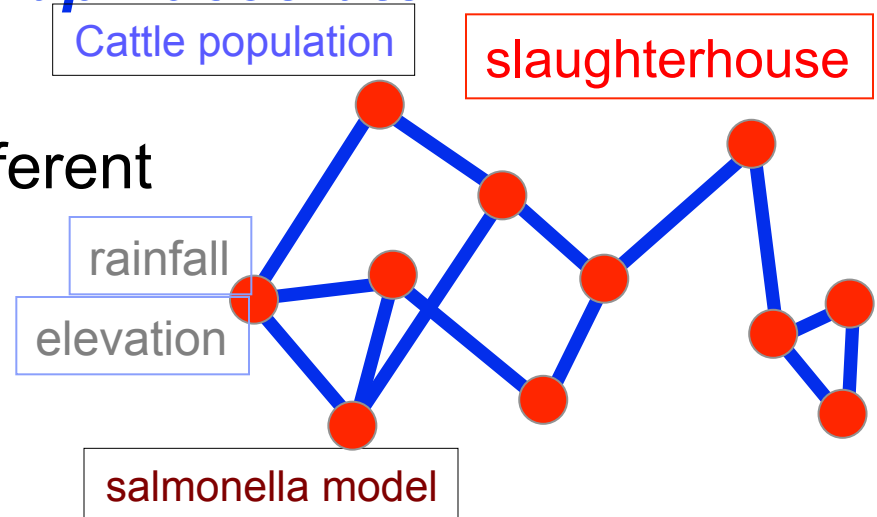# Foodborne Disease and Food Production
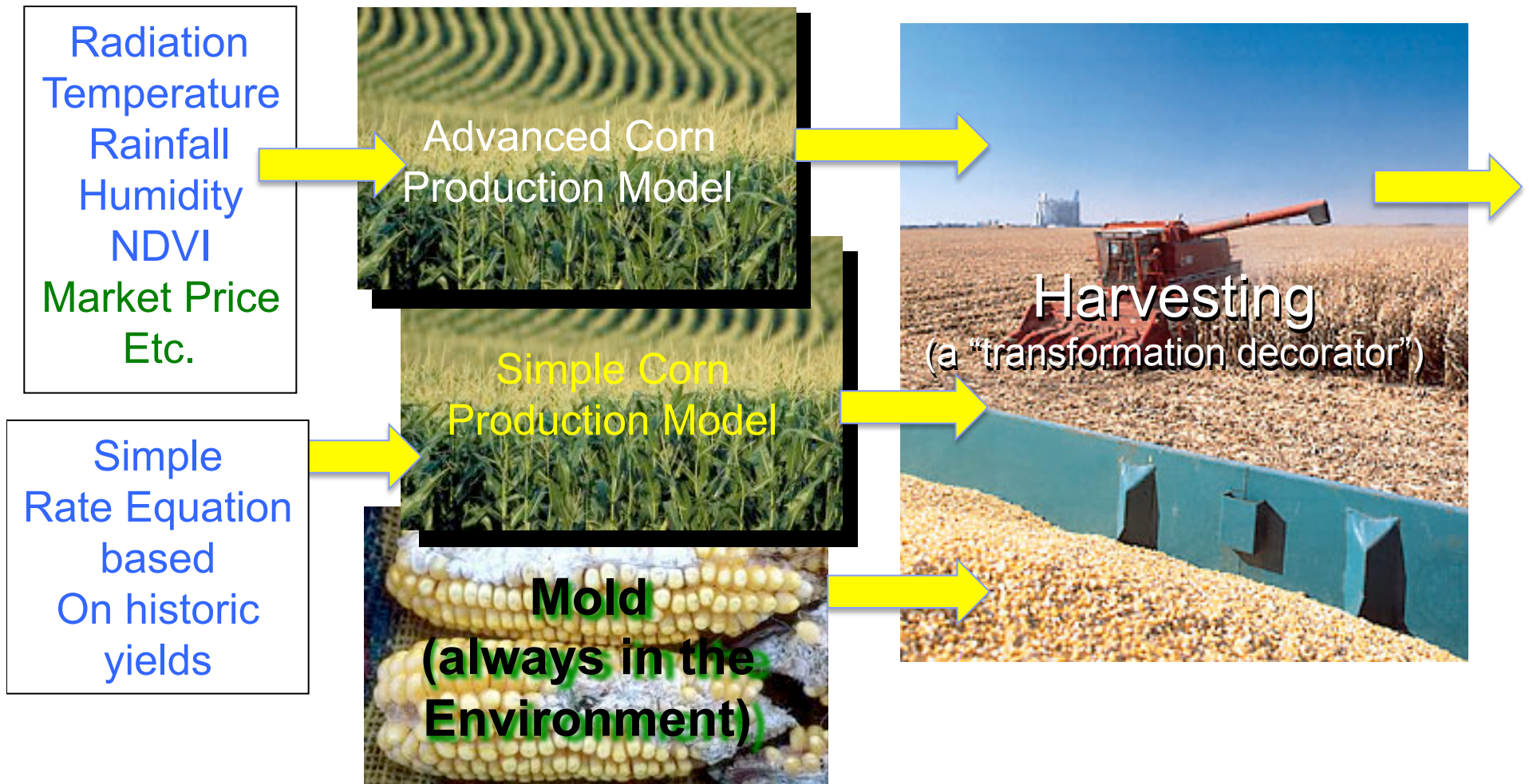## Modeling framework

with

Modeling food production requires a new kind of "*graph decorator*"

Nodes and Edges within a graph can be "decorated" with many different types of data and operators

- Labels
  - Static data
  - Time varying data
- Integration Decorators
  - Differential equations that can be solved
  - Disease models
  - Population models
  - etc.
- **Transformation Decorators**
  - Operators that can convert or map one graph object to another
  - Transformations may be discontinuous
  - Slaughterhouse
  - Contaminated slaughterhouse
  - Grain processing facility
  - etc.

Cattle population

slaughterhouse

rainfall

elevation

salmonella model

Gedanken experiment: Risk of Micotoxin in Corn

Radiation
Temperature
Rainfall
Humidity
NDVI
Market Price
Etc.

Simple
Rate Equation
based
On historic
yields

Advanced Corn
Production Model

Simple Corn
Production Model

Mold
(always in the
Environment)

Harvesting
(a "transformation decorator")

2. Corn Processing

Harvested Corn

| Residence time in Field | Storage time on farm | Controlled Storage | Time on Ship |

1. Mold Growth Model
2. Micotoxin Accumulation Model

Historic geocoded time series
Radiation, Temperature
Rainfall, Humidity

3. Receipt and Final Manufacturing
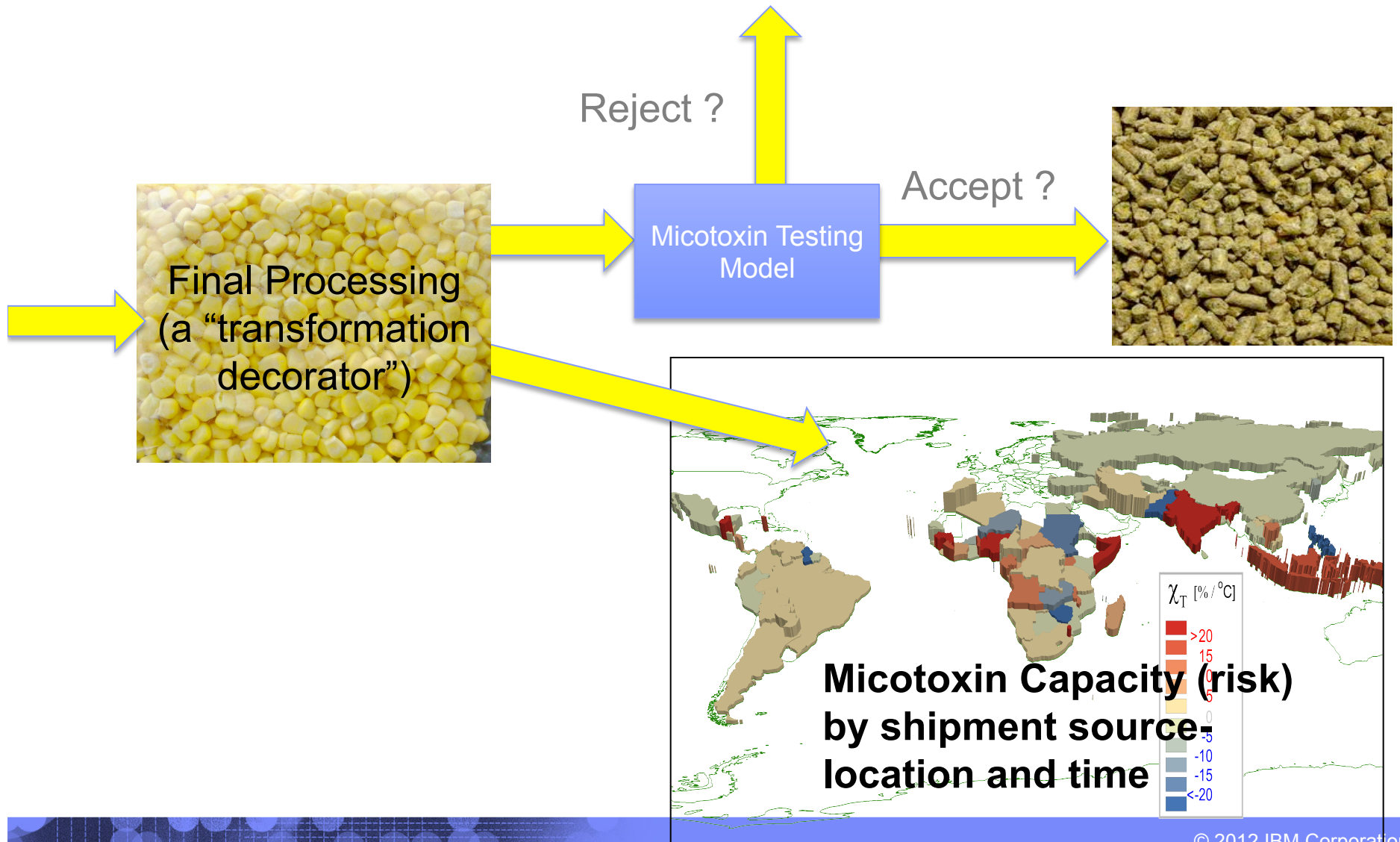
Final Processing (a "transformation decorator")

Reject ?

Accept ?

Micotoxin Testing Model

χ_T [% / °C]

>20
15
0
5
-5
-10
-15
<-20

Micotoxin Capacity (risk) by shipment source location and time
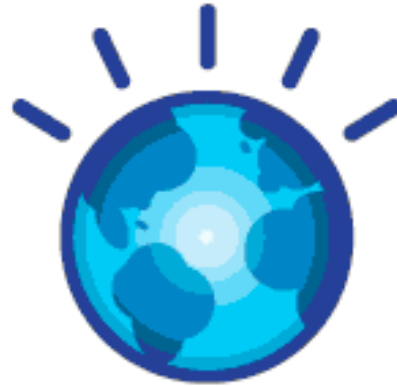
Proactive modeling of food borne disease outbreaks
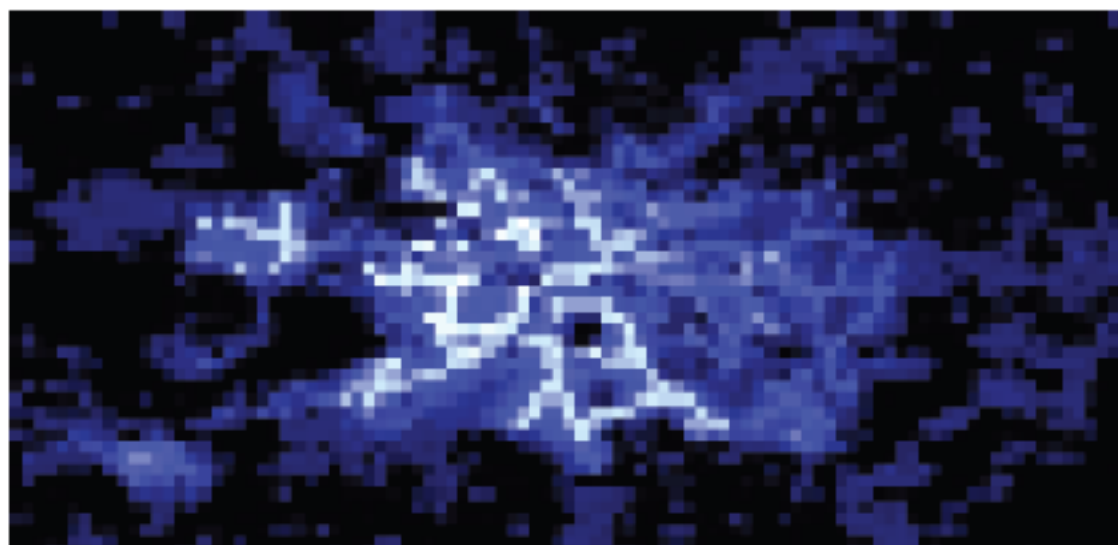
# Can we begin to investigate an outbreak before it even occurs ?

EHEC/STEC outbreak in Germany 2011

▶ Shiga-toxin producing *E. coli* bacteria

▶ Bacteria cause Hemolytic-uraemic syndrome (HUS)

▶ Symptoms: hypertension, bloody diarrhea, kidney failure

▶ investigation revealed that (fenugreek) sprout seeds were contaminated source

▶ 3091 reported cases in Germany, ca. 4000 in Europe

▶ 47 people died
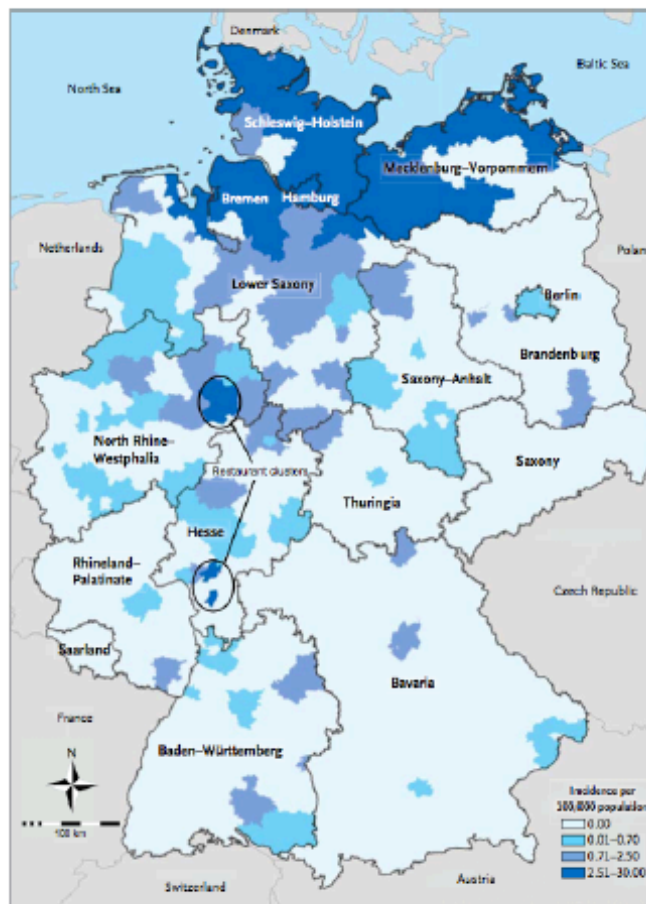
(source: Frank *et al.*, New Engl J Med, 2011; Ergebnisbericht der Task Force EHEC zur Aufklärung des EHEC O104:H4 Krankheitsausbruchs in Deutschland, 2011)
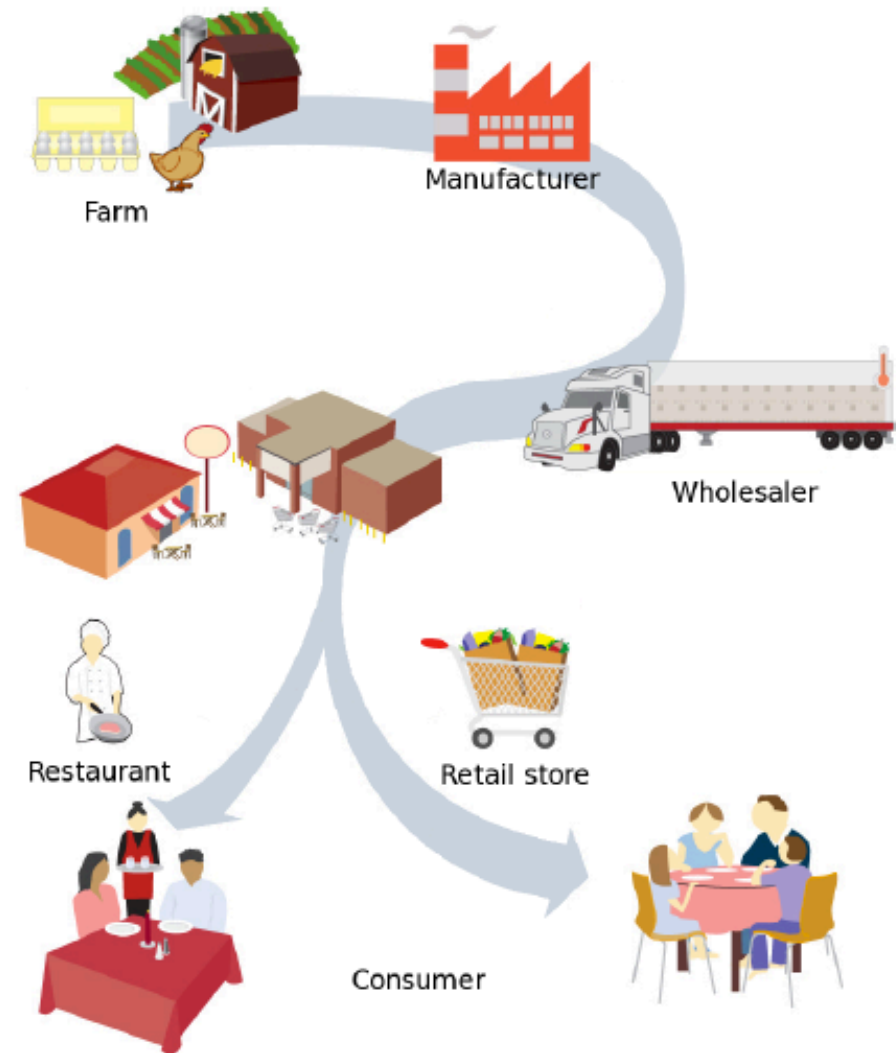
Epidemic curve
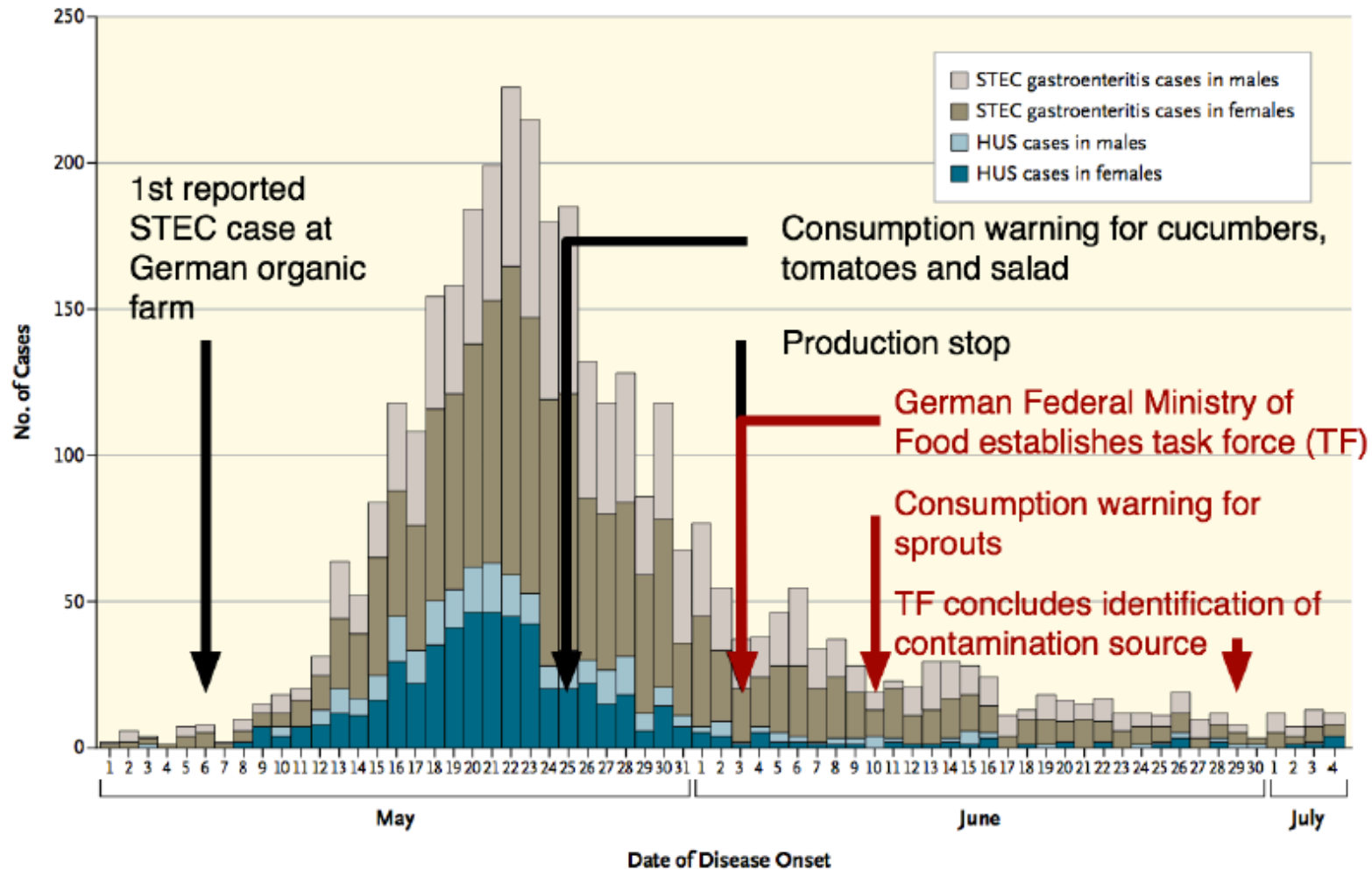
(source: Frank et al., New Engl J Med, 2011; Ergebnisbericht der Task Force EHEC zur Aufklärung des EHEC O104:H4 Krankheitsausbruchs in Deutschland, 2011)

From Farm to Ford: How do people get food?

Food supply chain:

1. (Farms, Manufacturers)
2. Wholesalers
3. Supply chain network
4. Retail/ grocery stores and restaurants
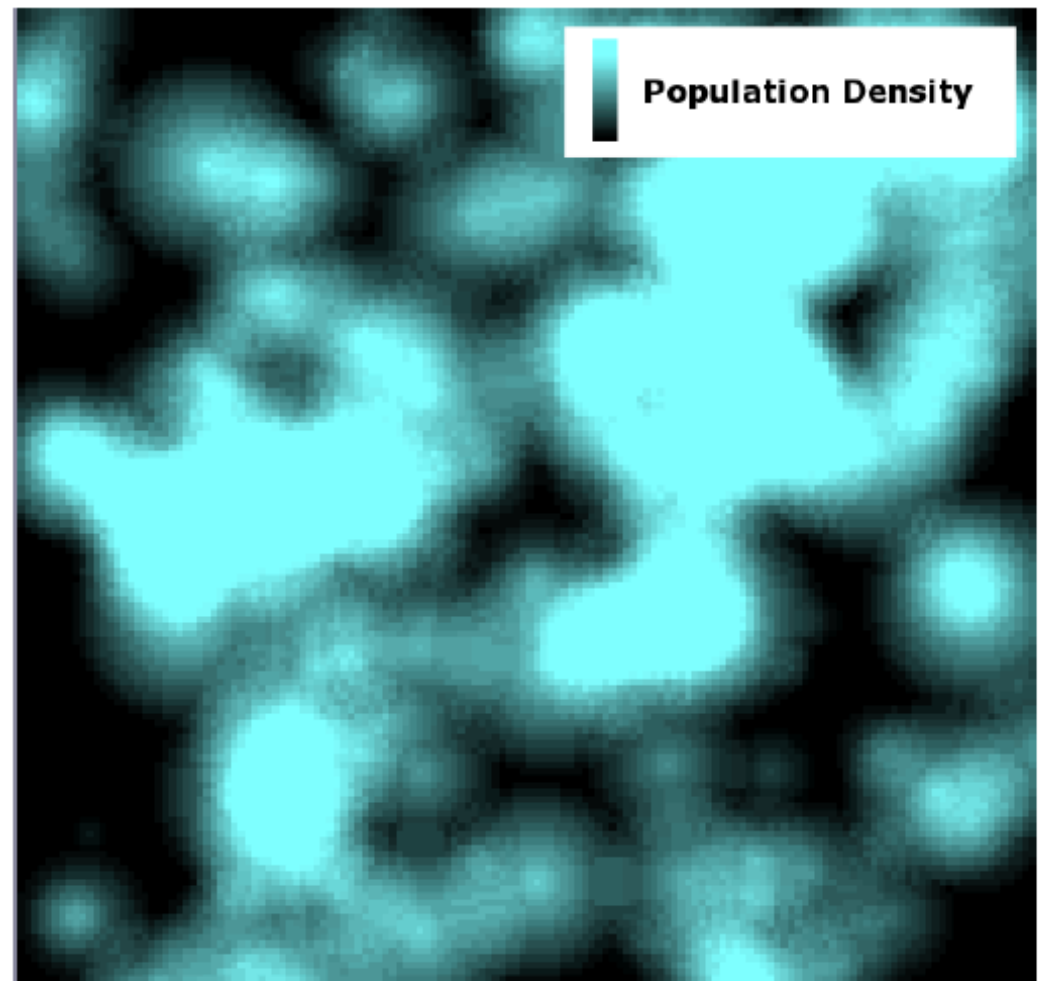5. Consumers

A Model World

Food Distribution Network:

Wholesalers
- Trade 1 type of food
- 1::n relation with retailers
- Scale-free network (power law distribution of retail stores. Degree = k)

People
- Toy model is a 2d grid
- Random population centers
- Density falls off linearly

Population Density

Retail stores and consumer behavior

Retail stores

▶ Uniformly distributed

▶ Huff's gravity model

  ▶ A_j – attraction factor of retail store j

  ▶ d_{ij} – distance between location i and retail store j
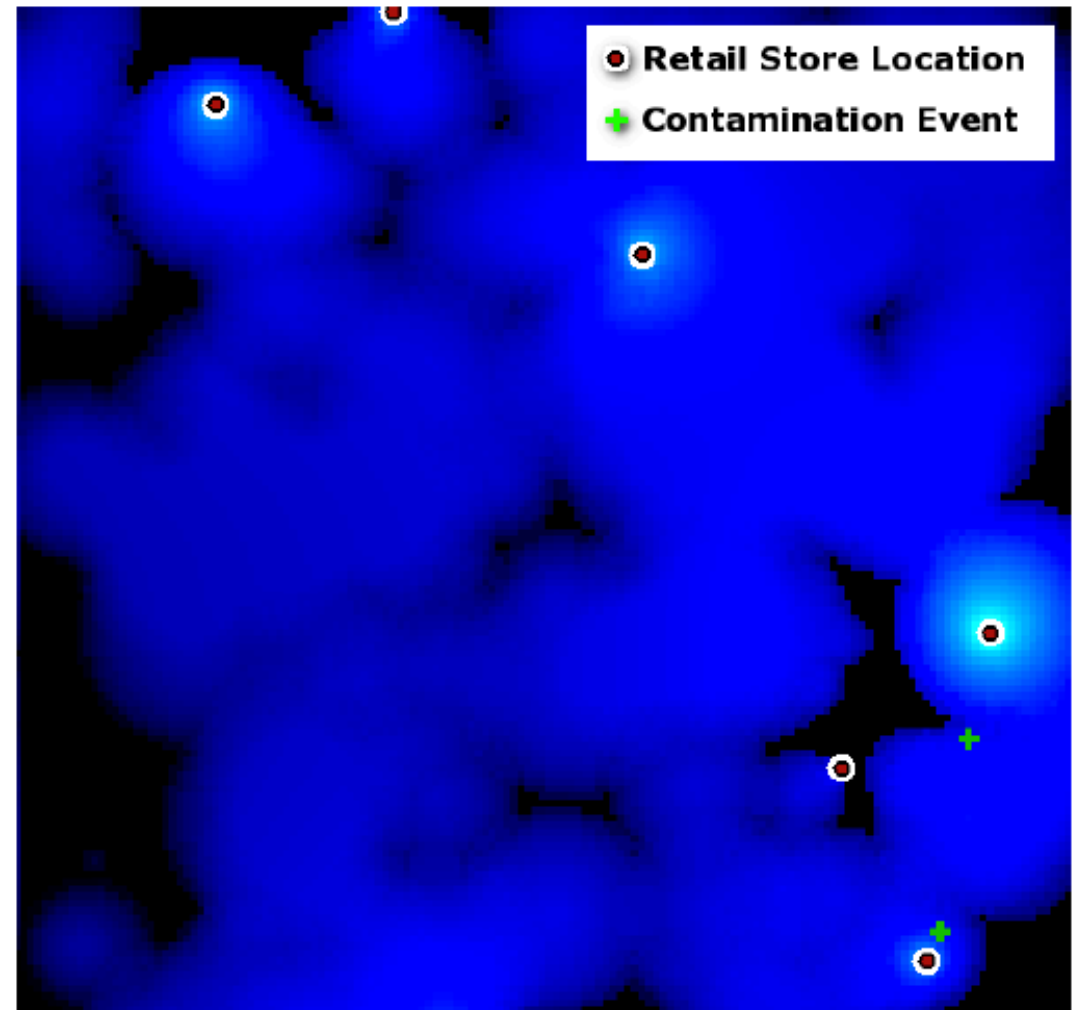
  ▶ γ – distance decay rate

  ▶ $p_{ij} = \dfrac{A_j \cdot d_{ij}^{-\gamma}}{\sum_j A_j \cdot d_{ij}^{-\gamma}}$

Retail Store Location

Consumption Events

Food Distribution by *wholesale source* !!

Contaminations: sampling food consumption events

▶ One wholesaler affected

▶ Small percentage of food contaminated

Legend:
- ⊙ Retail Store Location
- ✛ Contamination Event

Which wholesaler distributed the contaminated food ??
Method: Maximum Likelihood Estimation (MLE)

Let $\theta = \langle 0, \ldots, 0, 1, 0, \ldots, 0 \rangle$ be a vector where the jth entry denotes the likelihood that a distributor j's food is contaminated:

$$
\begin{aligned}
L(\theta; i \text{ is infected and } i \text{ lives at } x_i, y_i) &= \mathbb{P}(i \text{ is infected and } i \text{ lives at } x_i, y_i \mid \theta) \\
&= \mathbb{P}(i \text{ lives at } x_i, y_i) \cdot \mathbb{P}(i \text{ is infected} \mid i \text{ lives at } x_i, y_i, \theta) \\
&= \varphi_{x_i, y_i} \prod_j [\mathbb{P}(i \text{ bought from } j \mid i \text{ lives at } x_i, y_i)]^{\theta_j}
\end{aligned}
$$

φ_{x_i, y_i} denotes the population density of i's location x_i, y_i.

Each distributor j is associated with a set of retail stores R_j:

$$
L(\theta; i \text{ is infected and } i \text{ lives at } x_i, y_i) = \varphi_{x_i, y_i} \prod_j \left[\sum_{k \in R_j} P(i \text{ bought at } k \mid i \text{ lives at } x, y) \right]^{\theta_j}
$$

Joint probability of a set of reported cases $\mathcal{D}$:

$$
L(\underset{\sim}{\theta}, \mathcal{D}) = \prod_{i \in \mathcal{D}} \varphi_{x_i, y_i} \prod_j \left[\sum_{k \in R_j} P(i \text{ bought at } k \mid i \text{ lives at } x_i, y_i) \right]^{\theta_j}
$$

Procedure

- ► Population grid size: 100×100
- ► 5 distributors
- ► 100 retail stores
- ► 100 experimental runs per data point, each time resampling
  - ► population
  - ► retail store locations
  - ► distributor/retailer network
- ► Contamination: sample consumption events according to the wholesaler's food distribution
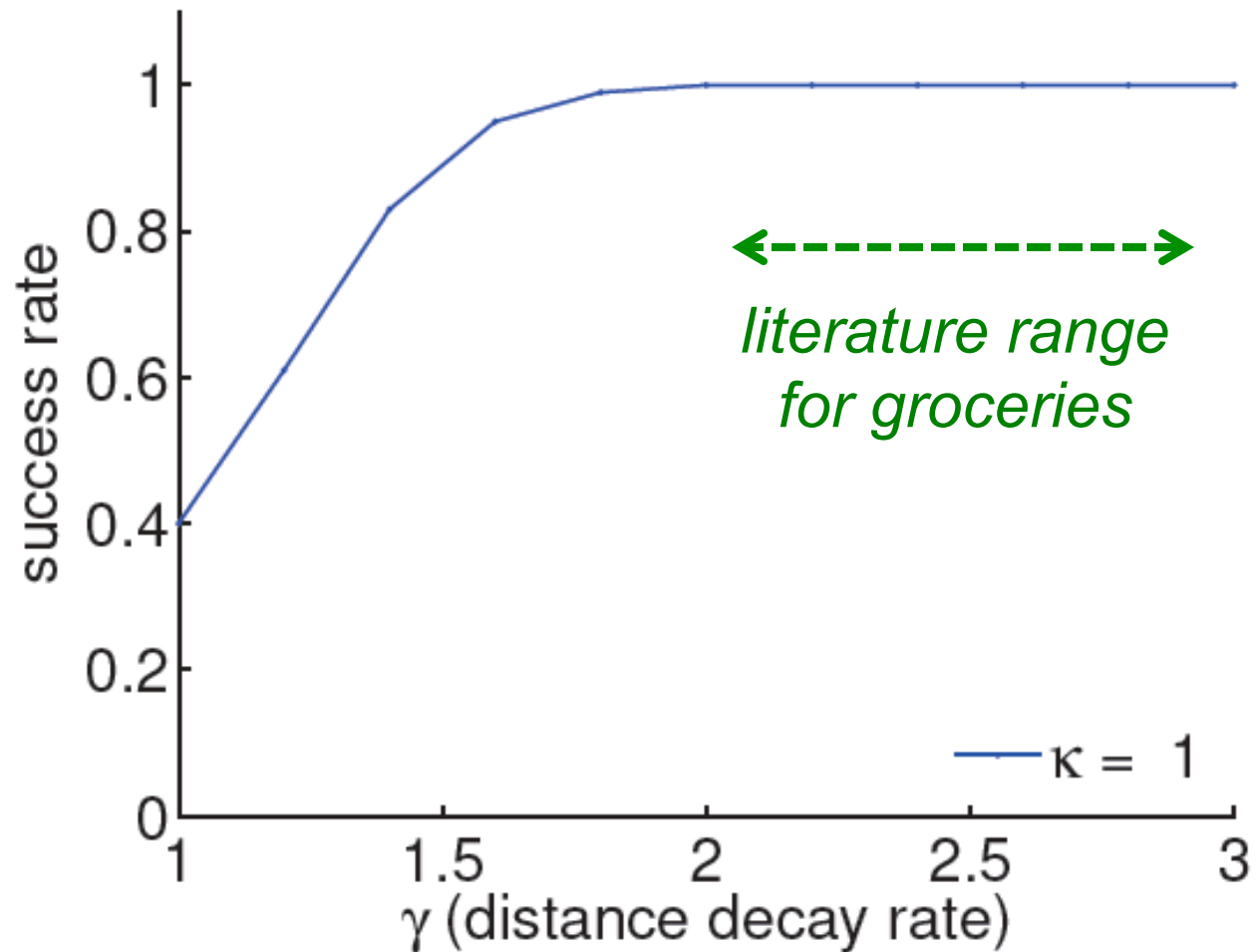- ► Every reported case is "correct"

Convergence of Method

distance decay exponent $\gamma = \{1, 1.6, 2.2, 2.8\}$
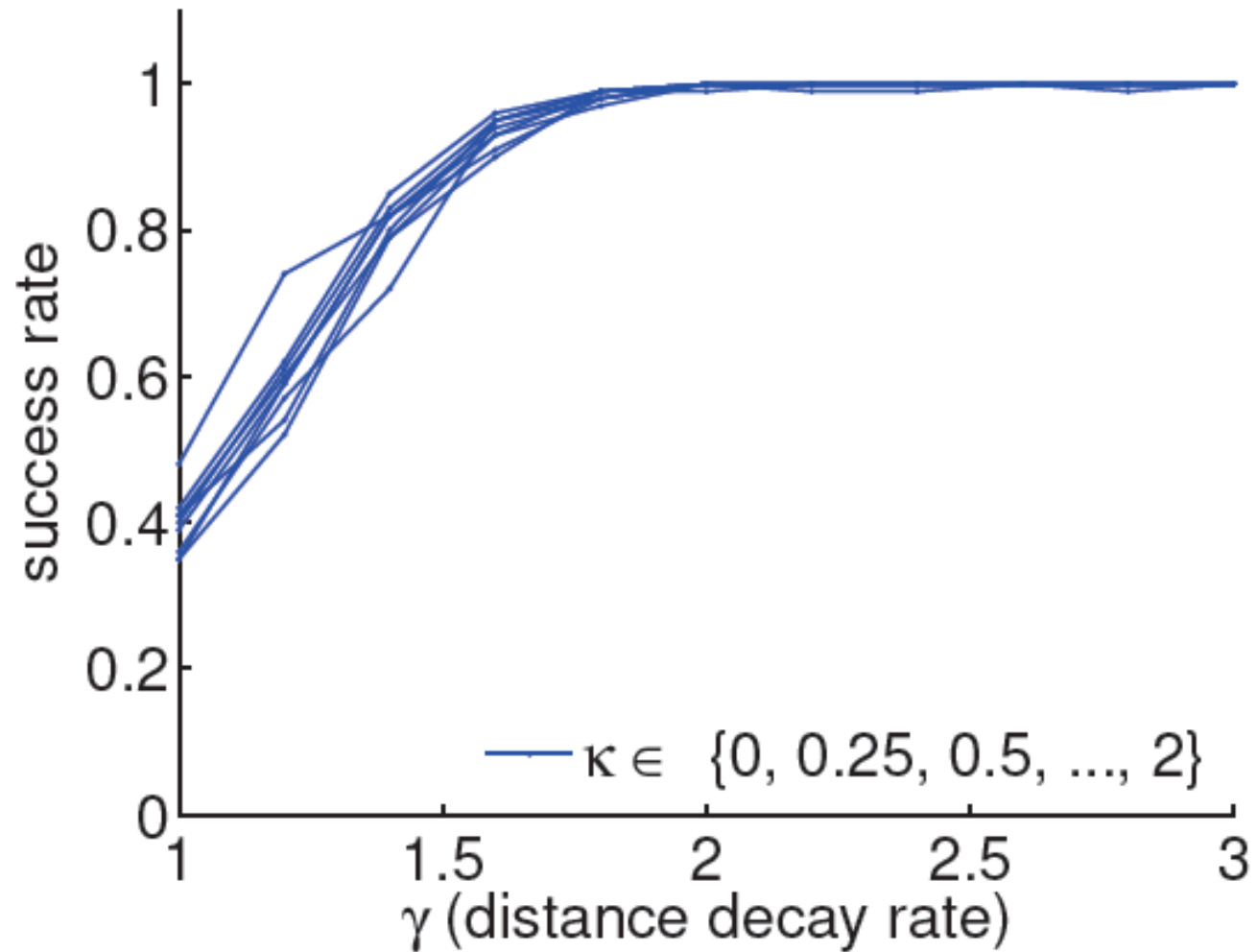   *literature range 2.0<γ<2.8*
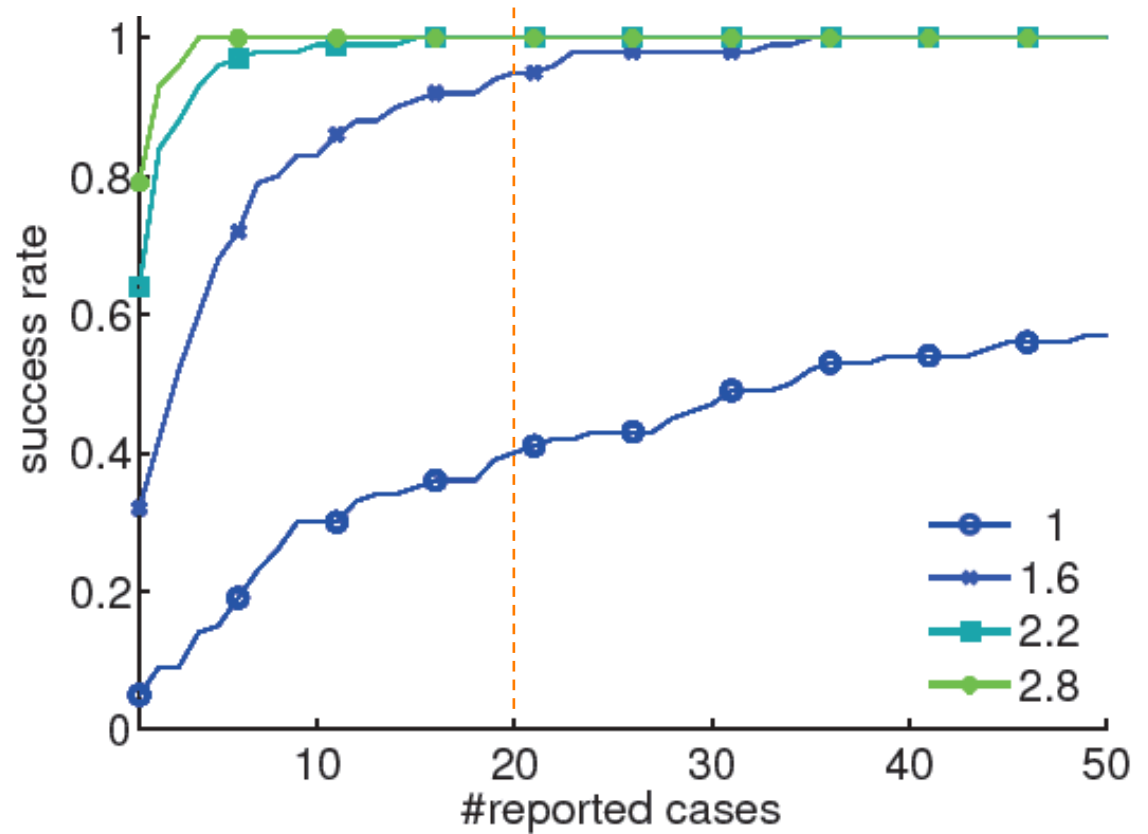
Dependence on Distance Decay Exponent

Number of reported cases: 20
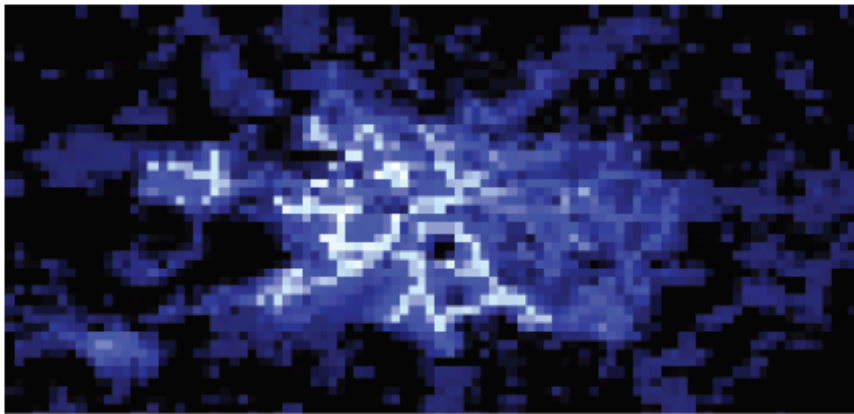
No strong dependence on network degree !!
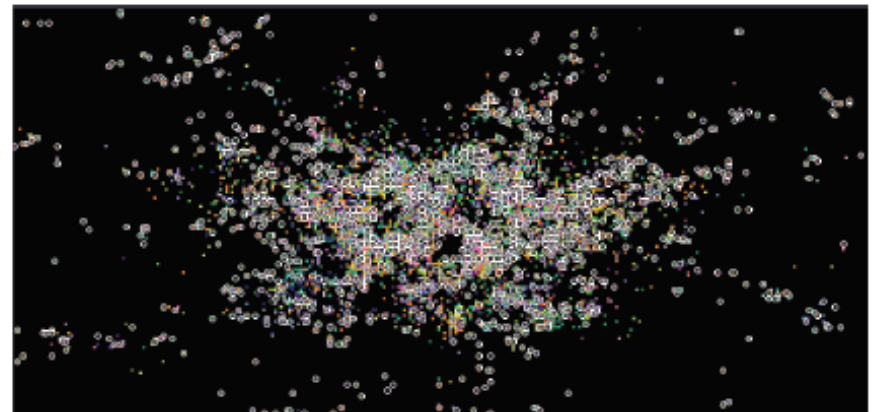

Number of reported cases: 20

next….

Berlin

Landscan census data from 2007

Retail locations from Google Places

- ▶ Cell size: $\sim 0.463km$
- ▶ Overall population size: $3,924,739$

- ▶ 1822 grocery stores
- ▶ 158 Lidl's, 138 Aldi's, 87 REWE's

August 2012

Next Steps

Enhancing the model:

- ▶ Extend model to several types of food
  - ▶ Incorporate pathogen-specific information
  - ▶ Outbreak characteristics
- ▶ Incorporate timely data
  - ▶ Harvesting periods of farmers
  - ▶ Production cycles of manufacturers
  - ▶ Delivery periods of wholesalers
- ▶ Study contamination during transport

Enhancing the prediction of affected wholesalers:

- ▶ Eliminate candidates (wholesalers) as contamination source
- ▶ Evaluate robustness of ML method against noise (false observations)

Enhance experiments:

- ▶ Incorporate empirical data from wholesaler distribution networks
- ▶ Compare simulations with data from real outbreaks