A SpatioTemporal Model for Influenza

Stefan Edlund¹, Michal Bromberg², Gabriel Chodick³, Judith Douglas¹, Daniel Ford¹, Zalman Kaufman², Justin Lessler⁴, Rachel Marom³, Yossi Mesika⁵, Roni Ram⁵, Varda Shalev², and James Kaufman¹

IBM Almaden Research Center¹, Israel Center for Disease Control (ICDC)², Maccabi Health Care Services³, John Hopkins School of Public Health⁴, IBM Haifa Research Center⁵

This project is being developed under Contract Number FA7014-07-C-0004, with the U.S. Air Force Surgeon General's Office (AF/SG) and administered by the Air Force District of Washington (AFDW).

August 2009
Talk Overview

- **Our task**
  - Disease modeling, data, sources of errors in the data, estimating model accuracy

- **Our solution**
  - Compartmental disease model with forcing function
  - STEM and experiments

- **Initial results**

- **Report on recent results using model optimization techniques**

- **Future work**
The problem at hand

- **Given 10 years worth of influenza incidence reports**
  - Build a model for influenza and fit the model to the reference data
    - Several options: Agent-based models, compartmental models, stochastic Markov-chains
  - Determine how quickly the model lose accuracy
    - Need a method to compare the result of model against the actual reference data
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
Sources of errors in the data

Misdiagnosis, under-reporting
Attacking the problem

- **We picked a compartmental disease model. Why?**
  - Commonly used in epidemiology and deeply studied
  - Models can be implemented and simulated efficiently
  - Useful when “law of large numbers” applies
    - Agent-based models useful for small scenarios tracking individual cases
  - Open source software exists to implement the model
    - Spatio-Temporal Epidemiological Modeler (STEM)
- **What is a compartmental disease model?**
Epidemiological Compartment Models

Standard SIR model

- Births
- $\mu^*$
- $\alpha$
- $\beta I$
- $\mu$
- $S \rightarrow I \rightarrow R$
- Deaths

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

\[
\begin{align*}
\frac{dS(t)}{dt} &= \mu^* T S(t) - \beta \frac{S(t) I(t)}{P} - \mu S(t) \\
\frac{dI(t)}{dt} &= \beta \frac{S(t) I(t)}{P} - (\gamma + \mu) I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t) - \mu R(t)
\end{align*}
\]

Standard SEIR model

- Births
- $\mu^*$
- $\alpha$
- $\beta I$
- $\varphi$
- $\mu$
- $S \rightarrow E \rightarrow I \rightarrow R$
- Deaths

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

\[
\begin{align*}
\frac{dS(t)}{dt} &= \mu^* T S(t) - \beta \frac{S(t) I(t)}{P} - \mu S(t) \\
\frac{dE(t)}{dt} &= \beta \frac{S(t) I(t)}{P} - \varphi E(t) - \mu E(t) \\
\frac{dI(t)}{dt} &= \varphi E(t) - (\gamma + \mu) I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t) - \mu R(t)
\end{align*}
\]

The software

- **The Spatio Temporal Epidemiological Modeler (STEM)**
  - Available for free at [http://www.eclipse.org/stem](http://www.eclipse.org/stem)
  - Licensed under the Eclipse Open License

- **Collaboration by**
  - IBM research in California and Haifa, Israel
  - Universities: John Hopkins School of Public Health, University of Vermont and others
STEM implements the latest Java™ based Component Software Architecture

Every STEM component is a “Plug-in” or “bundle” that can be independently developed, deployed, used, extended, and shared.

- Existing plugins define geography, transportation systems, and population for the 244 countries and dependent areas.

The STEM framework comes from the Eclipse Foundation (Equinox):
- Based on industry-standard (OSGi) Software architecture
- Makes it possible to easily build one model on top of another

Users can:
- create their own models for a country or region
- build on existing models and create new ones, making STEM extensible, flexible, and re-usable.
STEM

- Comes pre-built with common compartment models
  - SI, SIR, SEIR and stochastic variants
- Models have a spatial component:
  - The number of new infections depend on the number of infected individuals in neighboring regions.

\[ \Delta I_j \propto \frac{\beta}{P_j} S_j I_j + \sum_k \frac{\beta}{P_j} \frac{m_{jk} P_j}{P_j + P_k} S_j I_k + \frac{\beta}{P_k} \frac{m_{jk} P_k}{P_j + P_k} S_j I_k \]
STEM Core Integration Engine

Choice Between

- Finite Difference Solver (fast, good for demos)
- Runge-Kutta-Fehlberg (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

- User Contributed Solvers
Influenza model

- We extended STEM with a new model
  - A “forcing” SIR disease model that captures the seasonality of the flu.

\[
\beta(t) = \beta_o \left[ (1.0 - \alpha) + \alpha \left| \sin(\Omega t + \phi) \right|^{\lambda} \right]
\]

Transmission rate function
Evaluating the model

- STEM comes with built-in tools to compare models
- First, we need to transform the Israel data to the STEM format
  - Comma separated files where columns are regions and rows are time (days):
    - 3 files:
      - S.csv (# of susceptible people)
      - I.csv (# of infectious people)
      - R.csv (# of recovered people)
Data preparation

- S, I, R was calculated by summing finite difference using incidence:

\[ \Delta S = -i(t) + \alpha R(t) + \mu (P - S(t)) \]
\[ \Delta I = i(t) - \gamma I(t) - \mu I(t) \]
\[ \Delta R = \gamma I(t) - \alpha R(t) - \mu R(t) \]

- Assumptions:
  - Recovery rate \( \gamma \) was 0.1 (ten day infectious period)
  - Immunity loss rate \( \alpha \) was 0.001 (immunity lasing approx. 3 years)
  - Birth and death rate from actual population data for Israel
  - Initial recovered population about 65 %
  - From estimates that in a given year about 30-40 % of people are susceptible to the seasonal flu.
Error function

- For each location and at every time step, calculate a root mean square error (RMSE) between the model and the reference data
  - In this report, we only looked at the $I$ (infectious) data
- Find a “best fit” by minimizing RMSE over time

$$RMSE(I_s, I_r) = \frac{\sum_{t \in T} RMSE(I_s, I_r, t)}{|T|}$$
Initial results

- We manually ran simulation in STEM and varied 5 parameters:
  - $\beta$ (transmission rate), $m_{jk}$ (mixing rate between neighboring regions), $\lambda$ (modulation exponent), $a$ (modulation amplitude) and $\varphi$ (modulation phase shift).
  - Simplification: $m_{jk}$ was assumed identical for each region

- A large parameter space was examined
  - Very time consuming!

- Initially fitted only against the first two years of data
Initial results
need to explore a large phase space of parameters!!

$\beta = 0.41, \ m = 0.9, \ \lambda = 0.55, \ a = 0.6$ and $\phi = 1.87$

Well mixed population
Initial results

- We also fitted against all 10 years of available data

\[ \beta = 0.40, \ m = 0.9, \ \lambda = 0.55, \ a = 0.6 \text{ and } \varphi = 1.85 \]

I = A/H3N2, II = A/H1N1 and III = B
Model accuracy loss over time

![Graph showing model accuracy loss over time](image)
Recent results

- A recent feature in STEM allows for automatically walking the parameter space and find an optimal set of parameters.
- Nelder-Mead algorithm.
Minimum at $\varphi = 1.75$, $\beta = 0.45$, $m = 0.9$, $\lambda = 5.7$ and $a = .34$
A single SIR(S) model: 2 year and 10 year fits to historic data

I = A/H3N2, II = A/H1N1 and III = B
* Modified error function ignoring data points between flu seasons

- 2 year fit model
- 10 year fit model
Future work

- The model we used is a simplification
- The dominant Flu strain circulating each year is different
  - We need a true “multi-serotype” model in STEM to capture it
  - A fitted multi-serotype model would allow us to ask “what if” questions:
    • E.g. If the dominant strain next year is H1N1, how many cases can we expect?
    • Can be used as input to epidemiologists and public health officials to determine flu-shot ingredients.
- Some initial results on a single strain H3N2 model has been implemented and submitted to ACM TOMACS
- A true “global” model without geographic limitations will be needed
Goal: Evaluation of Public Health Policies

Multiple parallel scenarios identically initialized from current real world conditions and simulate each simultaneously forward in time.
STEM links

- Main Web Site: http://www.eclipse.org/stem
- YouTube tutorial: http://www.youtube.com/watch?v=LfiibQX4IFE
  - http://www.youtube.com/watch?v=3S5DbjCHsx4 (Spanish version)
Global Flu Model Animation
Questions ???
Why Open Source?
Need for an International Community

Eclipse Open Healthcare Framework

UPMC
Pitt
Stanford
MIT
U. Helsinki
ICDC

MECIDS

University of Vermont (UVM)

In discussion/meetings planned

Northrop Grumman
France
(Government, Universities)

UC Davis
University of Edinburgh
Seasonal Transmission Rate Fluctuation