electronic Journal of Health Informatics
http://www.ejhi.net
2010; Vol 6, No 1 (2011): e9

A SpatioTemporal Model for Influenza

Stefan Edlund[1], Michal Bromberg[2], Gabriel Chodick[3], Judith Douglas[1], Daniel Ford[1], Zalman Kaufman[2], Justin Lessler[4], Rachel Marom[3], Yossi Mesika[5], Roni Ram[5], Varda Shalev[3], James Kaufman[1]

[1]IBM Almaden Research Center, San Jose, California, USA

[2]Israel Center for Disease Control (ICDC), Tel Aviv, Israel

[3]Maccabi Health Care Services, Tel Aviv, Israel

[4]John Hopkins School of Public Health, Baltimore, Maryland, USA

[5]IBM Haifa Research Center, Haifa, Israel

Abstract

This paper describes an attempt to model seasonal influenza using the SpatioTemporal Epidemiological Modeler (STEM). Ten years of influenza data collected at 49 locations in Israel by the Israeli Center for Disease Control was used to fit the model, and a deterministic SIR(S) compartmental disease model was extended to account for seasonal variation in transmission rates as well as mixing of infected individuals between geographic regions. An adaptive step-size ordinary differential equation solver generated the time series data, and model parameters were fitted to the first few seasonal cycles of the experimental data and compared to subsequent cycles. The model used a sinusoidal seasonal variation of the transmission rate using an exponent parameter, and the best fit exponent was found to be very close to the square root of a sine function (exponent = 0.55), rather than the simple sine function typically used. Other results found that the transmission coefficient was high most of the year and that spatial mixing was high with 90% of people regularly visiting adjacent regions. The study obtained an excellent fit for the first two annual cycles (with less than 3% Root Mean Square error), with rapidly degrading accuracy in subsequent years. Finally, the annual variation in influenza species and strain suggests that a multi-serotype model for seasonal flu should provide better predictive capability. STEM and the mathematical models used in the study are all open source, available at www.eclipse.org/stem. An extension of the model to incorporate multiple serotypes will be studied in the future.

Keywords: Simulation, Modeling, Health, Seasonal Influenza, Compartmental Disease Models

1. Introduction

Influenza epidemics occur every year, generally peaking in the winter and virtually disappearing during summer. The seasonal variation of influenza has been attributed to a number of possible factors including temperature, relative humidity, indoor crowding as well as variations in host immune response [1]. Also, since the influenza virus is constantly adapting (though a process called "antigenic drift"), immunity is not permanent as it is with some other viral diseases such as chicken pox. An individual is usually at least partially susceptible to new variants of the influenza virus within a few years.

Models for influenza and similar viral diseases have been extensively studied in the literature [2-6]. Recent experiments using guinea pigs seem to suggest that seasonal variation in influenza incidence can be explained by small changes in transmission rate triggered by changes in temperature and relative humidity [1]. Dushoff et al. [7] argue that – since SIR(S) disease models have an intrinsic

The *electronic Journal of Health Informatics* is an international journal committed to scholarly excellence and dedicated to the advancement of Health Informatics and information technology in healthcare. ISSN: 1446-4381
© Copyright of articles is retained by authors; originally published in the *electronic Journal of Health Informatics* (http://www.ejhi.net). This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License (http://creativecommons.org/licenses/by-nc-sa/2.5/au).

tendency to oscillate (although with dampened effect) – such changes can be further amplified by dynamic resonance caused by the population dynamics of the host-pathogen system. In this paper, we use this natural frequency when estimating a starting point for finding optimal disease model parameters.

Table 1: The Basic SIR(S)* Model

***Susceptible/Infectious/Recovered(Susceptible)**

$$\frac{S(t)}{dt} = -\beta(\frac{S(t)}{P})I(t) + \alpha R(t) + \mu(P - S(t))$$

$$\frac{dI(t)}{dt} = \beta(\frac{S(t)}{P})I(t) - \gamma I(t) - \mu I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t) - \alpha R(t) - \mu R(t)$$

***Table 1:** The Basic SIR(S)* Model*
**Susceptible/Infectious/Recovered(Susceptible)*

Our goal in this study is to test our ability to model any time series and to measure the rate at which models lose accuracy against that data. Given that the reference data and the simulations themselves are subject to approximations and experimental uncertainty, our focus is to identify sources of error and validate approaches to optimizing epidemiological models.

2. Methods

Influenza falls into a category of diseases that can be studied with a compartment model such as an SIR(S) model, representing the passage of individuals between the three states; susceptible, infectious, recovered (alternately referred to as removed), and eventually susceptible again. The model contains a bilinear interaction term that describes how infectious people transmit an infection to susceptible people as a simple product, $(\beta / P)\, SI$.

Three differential equations that define the time evolution of an SIR(S) model are shown in Table 1. Differential equations with only one independent variable (t) are called Ordinary Differential Equations (ODEs). Integrating these equations to determine S(t), I(t) and R(t) can be done numerically by performing the computation at various time step lengths and then picking a time step short enough to minimize the computational error. In this analysis we use the *Runge Kutta Fehlberg* method that takes advantage of such an adaptive step-size algorithm.

These three equations are subject to several assumptions and limitations. The first assumption is that the model parameters (e.g., γ,α) are all constant in time. A simple bilinear model with constant model parameters will never produce seasonal oscillations or epidemics. To overcome this limitation, a "forcing term" is typically added to β which modulates the transmission coefficient with a periodic function. We use a forcing term that included an exponent, λ, to provide independent control of the peak transmission "seasonal duty cycle" expressed as:

$$\beta(t) = \beta_o[(1.0 - a) + a\,|\sin(\varpi' t + \varphi)|^{\lambda}]$$

<Equation 1>

The equations discussed above define the time evolution of an infectious disease, but they do not represent the geographic distribution of people or how they move about in space. The influenza data we hope to model is *spatiotemporal* and represents time series from 49 locations in the state of Israel. For our simulation, the details of population distribution and transportation can be an important factor in understanding the evolution of an infectious disease in space and time.

To model the underlying distribution and circulation of the population, we use a transportation model represented by Equation 2. Consider two locations, *j* and *k*. For individuals at location *j*, the change in the infectious population has three terms, on site infection, infection from visitors originating at neighbor site *k*, and infections from susceptible individuals visiting site *k*. If β=β(t) is the same at sites *j* and *k*, and if m_{jk} represents the circulation of individuals between sites *j,k* for each time step, then the equivalent spatiotemporal differential equation for infectious individuals become:

$$\Delta I_j \propto \frac{\beta}{P_j} S_j I_j + \sum_k \frac{\beta}{P_j}\frac{m_{jk}P_j}{P_j + P_k} S_j I_k + \frac{\beta}{P_k}\frac{m_{jk}P_k}{P_j + P_k} S_j I_k$$

<Equation 2>

where the sum is taken over *all* neighbors $k \neq j$ and the computation must be done for all sites *j*. Here we have made the implicit assumption that $m_{jk} = m_{kj}$ which is true for populations with no net migration. In this first attempt to model the Israeli data, we included only interactions between adjacent locations. In general, more connections can be added to model air travel, etc. [8]

2.1. Data

The Israel Center for Disease Control (ICDC), Ministry of Health, operates an ongoing influenza surveillance system that is based on various clinical and virological data sources. Maccabi Health Care Services, the second largest health maintenance organization in Israel that serves about 25% of Israeli population (approximately 1,700,000 members), provides the ICDC's surveillance system with daily data regarding visits of its members to outpatient clinics

due to various diagnoses, including influenza like illness (ILI). Each record includes the patient's area of residence. The method used to obtain the data was described previously [9]. For the purpose of this study, the burden of influenza was assessed by the number of visits due to ILI, as it appears to match fairly well to the influenza activity based on virological data gathered since 1996 by the network of sentinel community-based clinics operated by ICDC during the winter months. From this ongoing work we obtained 10 years of summarized (daily) case reports. The areas of residence that were collected by the ICDC were mapped to 49 "natural regions" in Israel. The natural regions are administrative divisions of the 15 Israeli sub districts. The division of the State of Israel into districts, sub districts and natural regions is published by the Central Bureau of Statistics [8]. The mapping process was based on a regions file that is updated by the Central Bureau of Statistics once a year. This file includes 1445 regions with their relation to natural area, sub district, and district as well as details on their municipality status, population, religion, etc.Each case report corresponds to the identification of "influenza like symptoms" by a primary care physician; as such, they are subject to a variety of errors such as misdiagnosis, under-reporting, etc. The actual incidence of disease is expected to be proportional to the number of case reports. The fraction of incidence that actually gets reported depends on the fraction of individuals with the flu that go to the doctor as well as the fraction that are correctly diagnosed.

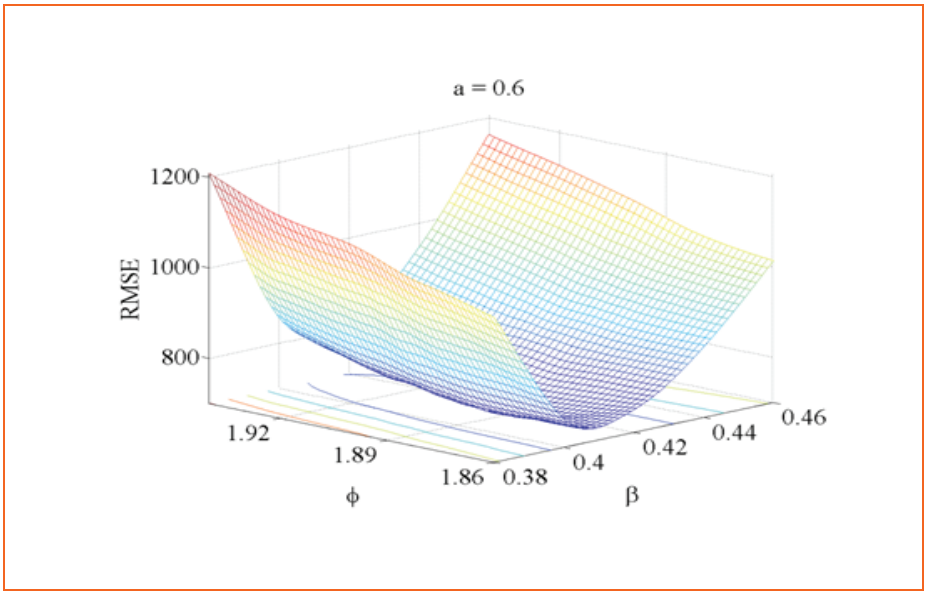

Figure 1a: *A minimum Root Mean Square Error (RMSE) at* $\varphi = 1.87$, $\beta = 0.41$, *a = 0.6. A much larger parameter space was studied to find the global minimum or best fit to the data.*

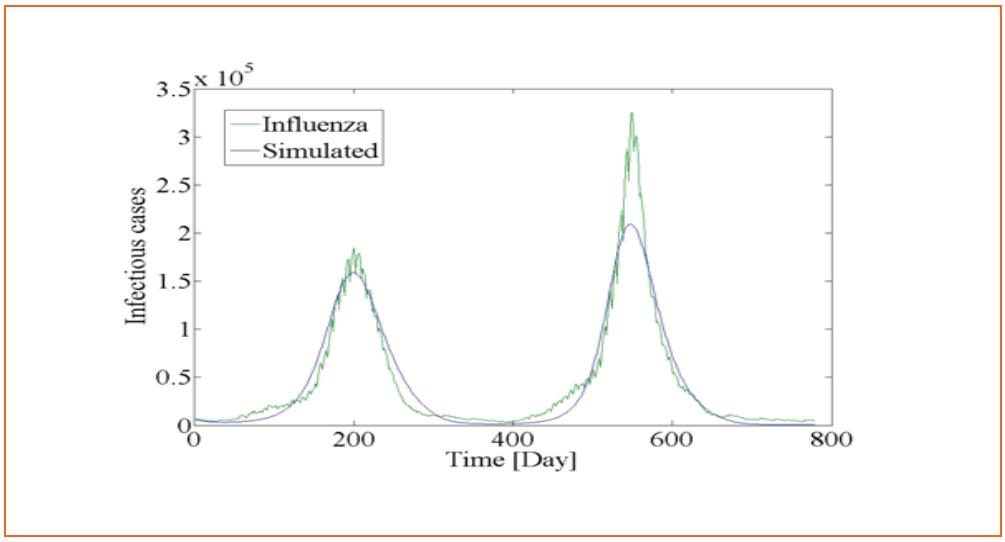

Figure 1b: *Infectious cases (summed over all locations) for fitted model (in blue) and actual influenza data (green).*

3. Results

To measure the accuracy of a model, it is desirable to estimate from the data the corresponding S(t), I(t) and R(t) values for the population. To estimate these values, we summed the finite difference equations for an SIR model using literature values for the recovery rate, $\gamma \sim 0.1$ (10 day infectious period), birth rates and death rates (from a linear population model for the state of Israel using actual population data), and the immunity loss rate, $\alpha \sim 0.001$ (immunity lasting about 3 years). The initial Recovered population and the reporting fraction were obtained using the constraints that the Susceptible population is positive definite, and that, on average, over long times, the derivative dS/dt should be near zero. Under these constraints, we found a "best" fit for an initial (and average) Susceptible population of 33.4% and an average reporting fraction of 1%. The reporting fraction was allowed to vary location by location to minimize drift in R and S over long times.

To optimize the parameters β, m, λ, a, and φ, we fit only the first two complete epidemic waves, and then studied how the model loses accuracy over the full 10 years of data. A large number of scenarios were studied varying the five model parameters over a wide range of values. Each scenario was initialized so that in each region the number of infectious cases as well as the initial number of immune cases matched the reference influenza data. The simulations ran 2 years of simulated time (returning a solution for each day) and a best fit was found that minimized the Root Mean Square Error (RMSE) measured against the actual influenza data. To estimate the overall "goodness" of a fit, we averaged the RMSE over time [9]:

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

<Equation 3>

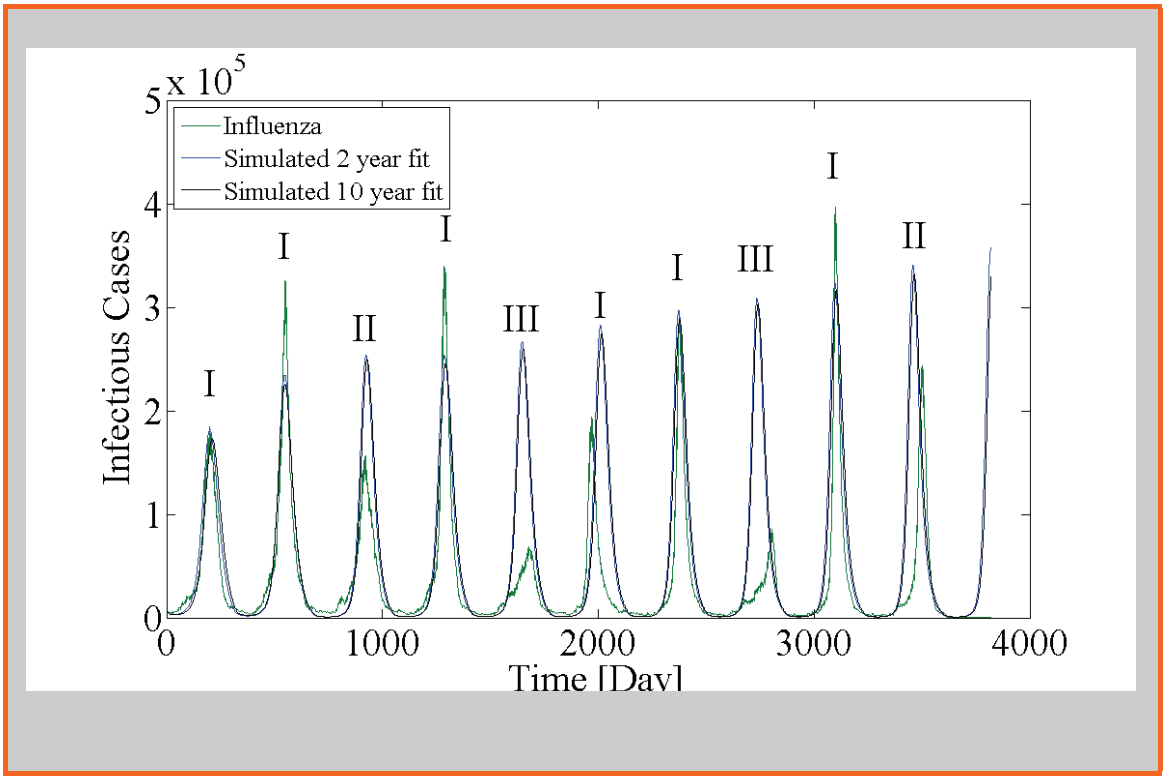

Figure 2: *Infectious cases as predicted by model over all 10 years of data for model fitted to the first two waves of the epidemic (blue) as well as all 10 years (black). The green curve is the reference influenza data. The roman numerals indicate the dominant influenza strain for the particular season: I = A/H3N2, II = A/H1N1 and III = B.*

A minimum RMSE was found using the following parameter values: $\beta = 0.41$, $m = 0.9$, $\lambda = 0.55$, $a = 0.6$ and $\varphi = 1.87$. Figure 1a shows the minimum for β and φ in a surface plot. Figure 1b shows the first 2 years of actual influenza cases (summed over locations) as well as the best fit model.

Using these same parameters we then ran the simulation for the full 10 years. As shown in Figure 2 (blue curve), the simulation loses accuracy beyond the second year and the rate of increase in RMSE (Figure 3, blue curve) clearly demonstrates the degradation in accuracy at longer times. In the first 2 years (the region used for the best fit), the largest source of error occurs in the wings of the epidemic wave. In the later years, the amplitude of the epidemic accumulates increasing the error.

4. Discussion

Our results were not unexpected. A seasonal SIR model does not capture the fact that the strain of influenza varies year to year. As shown in the Figure 2, the strain in year 3 was in fact different from the strain in years 1 and 2. Despite the limitations of using a single SIR model to study a multi-serotype infectious disease, it is still interesting to assess the process we used to develop a best fit predictive model. Fitting the five parameters based on the first two epidemic waves yielded a RMSE rate of increase of about 2.5% during the first two years increasing to about 4.5% at 10 years.

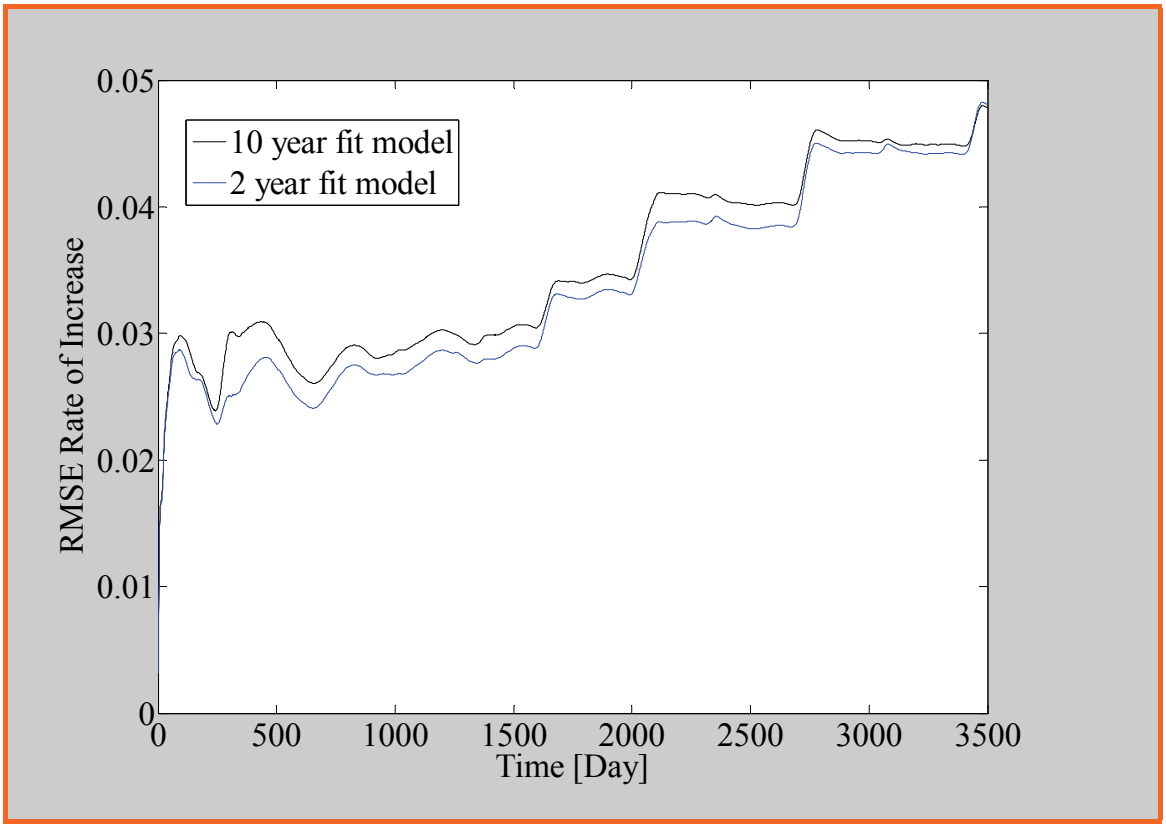

Figure 3: *Model accuracy loss as measured by the running RMSE rate of increase over 10 years. The blue curve is for the model fitted to the first two waves of the data only, while the black curve is fitted for all 10 years.*

Also shown in Figure 2 (in black) is a model fitted to all 10 waves of data ($\beta = 0.40$, $m = 0.9$, $\lambda = 0.55$, $a = 0.6$ and $\varphi = 1.85$). Interestingly enough the difference between fitting only the first two waves versus fitting all 10 years is very tiny. As shown in Figure 3, the 10 year model only becomes slightly better than the 2 year model towards the end of the 10 year period.

5. Conclusion

In this paper we demonstrated how a spatial SIR(S) model with a seasonal forcing function performs when used to model annual epidemics of influenza. Our forcing function introduced an exponent λ to allow control of the duty cycle for the transmission coefficient season modulation. An interesting finding from this SIR(S) model is that the best fit seasonal forcing exponent is very close to the square root of a sine function ($\lambda \sim 0.55$). Typically influenza models use a simple sine function (exponent $\lambda=1$). A square root dependence suggests that even in the subtropical climate of Israel, the transmission coefficient is high for more than half the year, suppressed only in the summer months. We also find that the model predicts a *basic reproductive number* R_0 for influenza of about 4.0, which is within the range found in other literature for a 100% susceptible population [7]. Our best fit to the spatio-temporal experimental data also demonstrates that for the state of Israel, the spatial mixing between regions is relatively high with 90% of people regularly visiting adjacent regions.

Finally, we conclude that an excellent fit can be obtained for the first two annual cycles with a rate of error increase of less than 3%. This is not surprising since the first two cycles were used to fit the model. However, in subsequent years the accuracy degrades rapidly. The annual variation in influenza species and strain clearly suggests that a multi-serotype model for seasonal flu should provide better predictive capability. A true multi-serotype model would track population resistance to common circulating strains and provide an estimate of the populations susceptibility to the strain expected in the next season (e.g., the dominant strain in circulation in the southern hemisphere) or, alternatively, predict the influenza strain(s) to which the population will be most susceptible in the next flu season. We plan to extend our model to incorporate multiple serotypes in future research.

Acknowledgements

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). The Air Force has not accepted the products depicted and issuance of a contract does not constitute Federal endorsement of the IBM Almaden Research Center. The authors acknowledge Eclipse for its support of the STEM as an open source project.

References

1. Lowen AC, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity and temperature. PLoS Pathog. 2007 Oct 25;3(10):e151. doi:10.1371/journal.ppat.0030151

2. Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintrau E, Bridges CB. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. Vaccine. 2007 Jun 28; 25(27):5086-96.

3. Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. New York: Oxford University Press; 1991.

4. Carrat F, Luong J, Lao H, Sallé A, Lajaunie C, Wackernagel H. A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics. BMC Med. 2006 Oct 23: 4:26. doi:10.1186/1741-7015-4-26. PMID 17059593.

5. Bootsma MC, Ferguson NM. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. PNAS. 2007 May 1: 104(18):7588-93.

6. Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. Nature. 2004 Dec 16: 432(7019):904-5.

7. Dushoff J, Plotkin JB, Levin SA, Earn JD. Dynamic resonance can account for seasonality of influenza epidemics. PNAS. 2004 Nov 19: 101(48):16915-6.

8. Lessler J, Kaufman JH, Ford DA, Douglas JV. The cost of simplifying air travel when modeling disease spread. PLoS ONE. 2009 Feb 6: 4(2): e4403. doi: 10.1371/journal.pone.004403.

9. Kaufman Z, Wong W, Peled-Leviatan T, Cohen E, Lavy C, Aharonowitz G, Dichtiar R, Bromberg M, Havkin O, Kokia E, Green M. Evaluation of a syndromic surveillance system using the WSARE algorithm for early detection of an unusual, localized summer outbreak of influenza B: implications for bioterrorism surveillance. Isr Med Assoc J. 2007;9(1):3-7.

10. http://gis.cbs.gov.il/shnaton59/all_israel2008.pdf and http://gis.cbs.gov.il/shnaton59/keymap_e.htm

11. Kaufman JH, Conant JL, Ford DA, Kirihata W, Jones B, Douglas JV. 2008. Assessing the accuracy of spatiotemporal epidemiological models. In: D. Zeng et al. (Eds): BioSecure 2008, LNCS 5354, pp. 143-154.

Correspondence

Stefan Edlund
Public Health and Computer Science Research
IBM Almaden Research Center
650 Harry Road
San Jose, California 95120-6099, USA

Phone: +0011 (1) 408 927 1766
http://www.eclipse.org/stem

edlund@almaden.ibm.com