Supporting information: Chaotic but regular recurrent influenza epidemics: from theory to observation

Sébastien Ballesteros[1,*], Lewi Stone[2], Anton Camacho[1]
Elisabeta Vergu[3], Bernard Cazelles[1,4]

[1]UMR 7625 (UPMC, ENS, AgroParisTech, CNRS), Ecole Normale Supérieure, Unit of Eco-Evolutionary Mathematics, 46 rue d'Ulm, F-75230 Paris Cedex 05, France.
[2]Biomathematics Unit, Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel
[3] INRA, UR341 Mathématiques et Informatique Appliquées, F-78352 Jouy en Josas, France
[4] UMMISCO UMI 209 IRD-UPMC, F-93142 Bondy, France.

[*] *Corresponding author*:
E-mail: sebastien.ballesteros@biologie.ens.fr

Contents

1 Models used **1**
1.1 A minimal model for two subtypes of influenza A interacting via a short period of full cross-immunity . . . 1
1.2 Stochastic metapopulation model with age structure . . . 3
1.3 Numerical Methods . . . 4

2 Parameter inference **5**
2.1 Methods . . . 5
2.2 Result for Ile de France data . . . 5
2.3 Result for Israel data . . . 8

3 Additional figures **9**
3.1 Single subtype model . . . 9
3.2 Two subtypes models . . . 11
3.3 Metapopulation model . . . 13

4 Periods bifurcations **17**

1 Models used

1.1 A minimal model for two subtypes of influenza A interacting via a short period of full cross-immunity

Figure 1 presents a schematic representation of equation 2 model for the 2 subtypes case.

Figure 1: A minimal model for two subtypes of influenza A interacting via a short period of full cross-immunity (Q) of average duration $1/q$. With the notation of eq 2 $J = \{\{\varnothing\}, \{1\}, \{2\}, \{1,2\}\}$ and trace the current level of immunity; $k = 1, 2$ and indexes the co-circulating subtypes. Infection (transition from states $R_{J\setminus k}$ to R_J) are represented by red and blue arrows for subtypes 1 and 2. During the transition hosts pass through states $I^k_{J\setminus k}$ and Q_J as denoted on the arrows. As the viral population evolves, hosts can loose immunity (black arrows).

1.2 Stochastic metapopulation model with age structure

Eq. 2 was extended to a stochastic metapopulation model of 52 cities depicted on figure 2 and indexed by c_l on the equations below.

Figure 2: Overview of the network of 52 global cities used in the stochastic metapopulation model. Circle areas and lines width are proportionate to city sizes and air transportation flows. Red, green and blue indicate the 3 geographic area used (North, Tropics, South).

$$
\begin{aligned}
\dot{R}_{\substack{J\\a_i,c_l}} &= qQ_{\substack{J\\a_i,c_l}} - \sum_{k\notin J}\sigma^k_{a_i}\beta_k(t)\frac{R_{\substack{J\\a_i,c_l}}}{p_{a_i}N_{c_l}}\sum_j M_{ij}I^k_{a_j,c_l} - \sum_{k\in J} g_k R_{\substack{J\\a_i,c_l}} + \sum_{k\notin J} g_k R_{\substack{J\cup k\\a_i,c_l}} \\
\dot{E}_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} &= \sigma^k_{a_i}\beta_k(t)\frac{R_{\substack{J\setminus k\\a_i,c_l}}}{p_{a_i}N_{c_l}}\sum_j M_{ij}I^k_{a_j,c_l} - 2\gamma E_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} - \sum_{m\neq l}\frac{p_{a_i}\tau_{c_l c_m}}{N_{c_l}}E_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} - \sum_{m\in J\setminus k} g_m E_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\notin J\setminus k} g_m E_{1\substack{(J\setminus k)\cup m\\a_i,c_l}}^{\;k} \\
\dot{E}_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} &= 2\gamma E_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} - 2\gamma E_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\neq l}\frac{p_{a_i}\tau_{c_m c_l}}{N_{c_l}}E_{1\substack{J\setminus k\\a_i,c_m}}^{\;k} - \sum_{m\neq l}\frac{p_{a_i}\tau_{c_l c_m}}{N_{c_l}}E_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} - \sum_{m\in J\setminus k} g_m E_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\notin J\setminus k} g_m E_{2\substack{(J\setminus k)\cup m\\a_i,c_l}}^{\;k} \\
\dot{I}_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} &= 2\gamma E_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} - 2\nu I_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\neq l}\frac{p_{a_i}\tau_{c_m c_l}}{N_{c_l}}E_{2\substack{J\setminus k\\a_i,c_m}}^{\;k} - \sum_{m\in J\setminus k} g_m I_{1\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\notin J\setminus k} g_m I_{1\substack{(J\setminus k)\cup m\\a_i,c_l}}^{\;k} \\
\dot{I}_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} &= 2\nu I_{1\substack{J:k\notin J\\a_i,c_l}}^{\;k} - 2\nu I_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} - \sum_{m\in J\setminus k} g_m I_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} + \sum_{m\notin J\setminus k} g_m I_{2\substack{(J\setminus k)\cup m\\a_i,c_l}}^{\;k} \\
\dot{Q}^k_{\substack{J\setminus k\\a_i,c_l}} &= \sum_{k\in J} 2\nu I_{2\substack{J\setminus k\\a_i,c_l}}^{\;k} - qQ_{\substack{J\\a_i,c_l}} - \sum_{k\in J} g_k Q_{\substack{J\\a_i,c_l}} + \sum_{k\notin J} g_k Q_{\substack{J\cup k\\a_i,c_l}}
\end{aligned}
$$

where $\frac{p_{a_i}\tau_{c_m c_l}}{N_{c_l}}$ are probabilities that a host of age class a_i situated in city c_l of size N_{c_l} travel from city c_l to city c_m (value given in (Grais et al., 2003)). 3 age classes (a_i) were retained with a_1 = 0-20 years old, a_2 = 20-60 years old and $a_3 > 60$ years old. Population sizes (N_{c_l}) and age distribution p_{a_i} were taken from https://www.cia.gov/library/publications/the-world-factbook/ and http://unstats.un.org/unsd/demographic/. To increase realism, exposed period was added (E) and exposed and infectious classes were doubled (E_1, E_2 and I_1, I_2) to ensure that the exposed and infectious period followed Erlang distributions. Following Cooper et al. (2006) it was assumed that only exposed individuals traveled.

1.3 Numerical Methods

Eq 1 and 2 were integrated by a Runge Kutta of order 4-5 with adaptive time step algortihm provided by the C library GSL (Galassi et al., 2003).

The stochastic metapopulation model was integrated using an Euler-multinomial approximation to the continuous-time Markov process written in C and using library GSL (Galassi et al., 2003) for random number generation and distribution.

2 Parameter inference

2.1 Methods

To compare the UPCA model to real data, we have implemented a stochastic version of eq 1 (adding a dephasing term d to the seasonal forcing function: $\beta(t) = \beta_0(1 + e\cos(2\pi t + d2\pi))$ using an Euler-multinomial approximation to the continuous-time Markov process.

We have contrasted the realisation of the process model (x_t) with weekly incidence data (y_t).

We have used 2 data sets: weekly incidence rates for Influenza Like Illness (ILI) of region Ile de France (Paris and its surrounding) from the french sentinelle network and Israel (LEWI, source). The French Sentinelle data are consultations for ILI in general practice. As the data come normalised per 100 000 hosts we have followed (Finkenstädt et al., 2005) and kept the population size fixed to 100 000 hosts. For Israel population size 1822438 source

Weekly incidence rates for the model prediction was computed as $x_t = \int_{t-1}^{t} \nu I(t)dt$ and contrasted to the data weekly reported incidence y_t with an observation process containing a reporting rate (ρ). We let $Y_t|\rho_t, x_t \sim P(\rho_t, x_t)$ where P stands for the Poisson distribution and, to account for overdispersion we allow variability in the reporting rate by assuming that $\rho_t \sim \Gamma(\frac{1}{\phi}, \rho\phi)$ (where Γ is the gamma distribution). The observation process is then fully described by a negative binomial distribution (NB): $Y_t|x_t \sim NB(\text{mean} = \rho x_t, \text{size} = \frac{1}{\phi})$.

Parameter inference was achieved with an implementation of maximum likelihood via iterated filtering (MIF) (Ionides et al., 2006; Breto et al., 2009) written in C using the standalone math library from R (R Development Core Team, 2008) for random number generation and distribution.

2.2 Result for Ile de France data

The results of the inference procedure is depicted in figure 3.

A strong correlation between β_0 and g was found as reported in figure 4.

Figure 3: Likelihood profiles (likelihood value following the evolution of a focal parameter, the others parameters being re-estimated by MIF) and likelihood slices (likelihood value fixing the value of all but the focal parameter at their maximum likelihood estimate) (inset). A stochastic version of the single subtype model with an observation process following a negative binomial distribution is fitted to the ILI french sentinelle network data of region ile de France. Two replications with different initial condition and algorithm parameters are plotted and local quadratic regression are used to estimate the profile likelihood.

Figure 4: Correlation between β_0 and g observed for two replications with different initial condition and algorithm parameters. Black: β_0 is fixed and the inferred value of g is plotted. Red: g is fixed and the inferred value of β_0 is plotted. In every cases, the other parameters are re-estimated to their maximum likelihood value.

2.3 Result for Israel data

Figure 5: likelihood slice

3 Additional figures

3.1 Single subtype model

Figure 6: Additional trajectories for the single subtype determinist UPCA model. Weekly incidence per 100 000 inhabitant are plotted. First line: $R_0 = 1.65$, $1/v = 2.47$ days^{-1}, $1/g = 7$ years^{-1}, $e = 0.104$, $\eta = 10^{-6.7}$; Second line: $R_0 = 1.65$, $1/v = 2.47$ days^{-1}, $1/g = 11$ years^{-1}, $e = 0.104$, $\eta = 10^{-7}$; Third line: $R_0 = 5$, $1/v = 8$ days^{-1}, $1/g = 20$ years^{-1}, $e = 0.35$, $\eta = 10^{-6}$; Fourth line: $R_0 = 2.66$, $1/v = 2.77$ days^{-1}, $1/g = 20$ years^{-1}, $e = 0.2$, $\eta = 10^{-7}$

3.2 Two subtypes models

Figure 7: Additional trajectories for two subtype determinist (first and second line) and stochastic metapopulation UPCA model (third line). Blue and red lines are weekly incidence per 100 000 inhabitant for subtype 1 and 2 and black lines sum of both. First line: $R_0 = 5$, $1/v = 8$ days^{-1}, $1/g = 20$ years^{-1}, $1/q = 6$ months^{-1}, $e = 0.35$, $\eta = 10^{-6}$; Second line: $R_0 = 2.66$, $1/v = 2.77$ days^{-1}, $1/g = 25$ years^{-1}, $1/q = 6$ months^{-1}, $e = 0.2$, $\eta = 10^{-7}$; Third line: Typical realisation for the city of Paris $R_0 = 2.66$, $1/v = 2.77$ days^{-1}, $1/\gamma = 1.5$ days^{-1}), $1/g = 25$ years^{-1}, $1/q = 6$ months^{-1}, $e = 0.2$

3.3 Metapopulation model

Figure 8: 20 realisations of the stochastic metapopulation model aggregated by geographic areas. Blue and red lines are weekly incidence per 100 000 inhabitant for subtype 1 and 2 and black lines sum of both; . Parameters are taken from the empirical parameter set with $1/g = 25$ years^{-1}, $1/q = 6$ months^{-1}, $1/\gamma = 1.5$ days^{-1}). Seasonal forcing ($e = 0.2$) for Northern and Southern hemisphere cities are in phase opposition and tropics do not present seasonality.

Figure 9: Per city view of a realisation of the stochastic metapopulation model.City names are color coded by geographic area with red for North, green for Tropics and blue for South. To facilitate visibility, each lines represent the ratio of weekly incidence per 100 000 inhabitant over its maximum value. Parameters are identical to figure 8.

Figure 10: Example of trajectory obtained in different cities for the 3 geographic areas (North, Tropics and South). Blue and red lines are weekly incidence per 100 000 inhabitant for subtype 1 and 2 and black lines sum of both. Parameters are identical to figure 8.

4 Periods bifurcations

Figure 11: Bifurcation diagram for the single subtype model of figure xxx. Two different measures of periodicity are plotted in the graphs: Measure 1: period is determined as the smallest T for which $|log(I(t+T) - log(I(t))| < \epsilon$ where ϵ is a small numerical tolerance. Measure 2: period is determined by the analysis of Fourier spectra. As the annual signature of seasonal forcing is always present, an arbitrary parameter ϵ' was used to consider that others period are insignificant. Left and right graphs are respectively for theoretical and empirical parameter set. Top and bottom graph are respectively for period as measured by measure 2 and 1.

Figure 12: Same legend as for figure 11 for the two subtypes model of figure xxx. The periodicity of I1+I2 is plotted.

Figure 13: Same legend as for figure 11 for the single subtype model of figure xxx. Left and right graphs are respectively period as measured by measure 2 and 1 (see caption of figure 11 for definition).

References

Breto, C., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. *ANNALS OF APPLIED STATISTICS*, 3:319.

Cooper, B. S., Pitman, R. J., Edmunds, W. J., and Gay, N. J. (2006). Delaying the international spread of pandemic influenza. *PLoS Med*, 3(6):e212.

Finkenstädt, B. F., Morton, A., and Rand, D. A. (2005). Modelling antigenic drift in weekly flu incidence. *Stat Med*, 24(22):3447–3461.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., and Rossi, F. (2003). *Gnu Scientific Library: Reference Manual*. Network Theory Ltd.

Grais, R. F., Ellis, J. H., and Glass, G. E. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur J Epidemiol*, 18(11):1065–1072.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proc Natl Acad Sci U S A*, 103(49):18438–18443.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.