Supplementary Information:

Jacob Bock Axelsen, Rami Yaari, Bryan Grenfell and Lewi Stone

May 12, 2014

1 Methods

1.1 Modelling Regular Outbreak Dynamics

The epidemics dynamics were modelled based on a deterministic age-of-infection SIR model [1], which we extended to an SIRS model by adding a simple feedback from the recovered population to the susceptible population via loss-of-immunity caused by steady antigenic drift. In this model the number of newly infected at time t is:

$$i(t) = \frac{S(t)}{N} R_0(1 + \delta(t)) \sum_{\tau=1}^{d} P_\tau i(t - \tau) \tag{1}$$

where S is the number of susceptible, R_0 is the basic reproduction rate of a fully susceptible population and P is the seven-day infectivity profile ($d = 7$) chosen as a Gamma distribution with a mean of 2.7 days and variance 1.8 days and N is the reporting population at a reporting rate of 10%. The dynamics of the susceptible pool is then:

$$S(t) = S(t - 1) - I(t) + \lambda R(t - 1) \tag{2}$$

Recall $1/\Lambda$ is the average period of immunity of a typical individual immunity usually 2 to 10 years so that Λ typically ranges .1 to .5, and for a model with daily time step, we set $\lambda = \Lambda/365$. The infected individuals recover after d days, thus the recovered population growth is:

$$R(t) = R(t - 1) + I(t - d) - \lambda R(t - 1) \tag{3}$$

Finally there is also a constant influx of influenza arriving in Tel Aviv from any destination in the world, which was set at one person per day all year round. The influenza influx cannot be modelled higher than 10 per day as this gives a model baseline ILI above one influenza case per day in the low season ($10 \times 10\% = 1$, taking into account a 10% reporting rate). Conversely, an influx rate much below 0.1 sick persons per day will eventually render the

model unable to fit to the 'small' outbreaks since the infected and susceptible population may often be unable to reach the epidemic threshold $R_{\text{eff}} = 1$ in those years.

The simple climate driver is modeled as $R_0 = \overline{R_0}[1 + \delta \sin(\omega t)]$. The real climate driver is expressed as a linear combination thus: $\delta(t) = w_{\text{T}}\text{T}(t) + w_{\text{RH}}\text{RH}(t)$ where T is daily average temperatures and RH is relative humidity, both normalized to the interval $[-0.5, 0.5]$.

1.2 Fitting Epochal Years

The outbreaks during the epochal jump years (2003, 2009) are different from the other years in both timing and dynamics. As emphasised in the paper, a key feature of the A-Fujian 2003 year when the epochal jump occurs, is that the epidemic initiates two months earlier than usual (i.e., as compared to other years). The 2009 H1N1 year also initiates early (in fact out of season) and its main peak is also two months earlier than usual. The full 2009 epidemic trajectory is obscured by multiple peaks in the data. Both the 2003 and 2009 outbreaks are very spiky and they do not conform to normal SIR modelling. This motivated fitting two extra parameters to be used only in years in which there are antigenic jumps.

The simplest assumption for a year with an epochal jump is the inclusion of fast. high loss of immunity from the beginning of the season when the new virus strain appears. Thus to keep things simple, if an epochal year is identified, immunity loss is immediate from the start of the season and continues for the whole year.

The other feature of the A-Fujian year when the epochal jump occurs is that the epidemic initiates two months earlier than usual. The only clear way known to the authors for initiating an early epidemic is to slightly increase R_{eff} (see Stone et al. (2007)). We thus fit an extra parameter to the model which allows for the possible extra increase in R_{eff} by adding a pulse over the first 6 months of the season when the epidemic grows.

The antigenic jump is modelled first as a modification to $\delta'(t) = \delta(t) + \Delta$, $\Delta = 0.26(0.02)$, for the period leading up to 1st December and a modification in immunity-loss rate $\lambda' = \lambda + \lambda^*$, $\lambda^* = 0.04(0.01)$, for the whole year.

2 The Bayesian inference: Parameter estimation and prediction accuracy

In the main text we reported a high correlation r=0.94 between the model ILI fit and the observed ILI data. In Fig. 1 we plot a 2D histogram demonstrating the success of the fit.

We assume that the six prior marginal parameters θ^i: basic reproduction rate R_0, relative humidity weight ω_{RH}, temperature weight ω_{T}, loss-

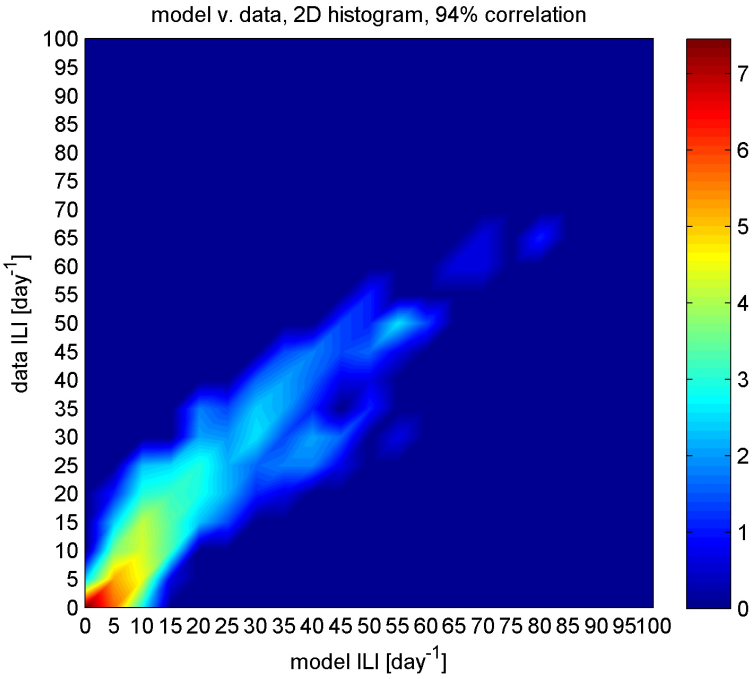

Figure 1: 2D histogram of daily ILI observed versus the model output ILI. The color bar is logarithmic.

of-immunity rate λ, seasonality driver enhancement Δ, loss-of-immunity enhancement λ^* are vaguely distributed according to broad Gamma distributions. The boundary condition on susceptible population is assumed uniformly distributed $S_0 = S(0) \sim U[0,1]$. The density functions of these marginal priors are $P_m(\theta^i)$.

We construct a prior likehood function from the residuals of the ILI counts $\hat{I}(t)$ and the model outcome $I(t) = M(t, R_0, \omega_{\mathrm{RH}}, \omega_{\mathrm{T}}, \lambda, \eta, \lambda^*)$. The residuals, $\Delta^\theta(t) = \hat{I}(t) - I(t)$, are assumed to be Gaussian distributed $\Delta(t) \sim N(0, \sigma^2)$ with probability density $P_N(\Delta^\theta(t))$. The prior likelihood is then $\log(Pr) = \sum_t \log(P_N(\Delta^\theta(t))$.

According to Bayes' theorem $p(\theta|\Delta) \sim p(\Delta|\theta)p(\theta)$ and thus the posterior log-likelihood function $\log(Ps)$ obeys:

$$\log(Ps) \sim \sum_t \log(P_N(\Delta^\theta(t))) + \sum_i^8 \log(P_m(\theta^i)) \tag{4}$$

The Monte Carlo Markov Chain sampling is then performed using the slice sampling method [2] on $\log(Ps)$. The sampling is continued until convergence on parameter values are attained and the output trace for each parameter displays minimal autocorrelation.

With regard to the priors, we followed the common procedure of using Gamma distributions for parameters that are positive. The shape/scale has basically no effect since the sampler is not restricted by the density of the pdf in the regions that provide the best fit. The hyper-parameters do not influence the procedure. We have used other distributions with little effect. The deeper understanding is that the parameter settles according to the data, rather than the mean and variance of the prior.

In Table 1 we have specified the priors and values of hyperparameters that were used for the Bayesian inference.

The software used was Matlab R2011b for 64-bit Windows 7 with the statistics toolbox using the slice sampling algorithm "slicesample" [2]. The procedure followed is standard bayesian MCMC analysis and validation methods which is documented in the manual to the "slicesample" function in Matlab.

In Fig. 2 we show the density kernels of the MCMC sampling procedure for the six parameters. Best fitting parameters (mean and 95% credible intervals): $R_0 = 2.9(\pm 0.1)$, $S_0 = 0.31(\pm 0.02)$, $w_T = 0.13(\pm 0.02)$, $w_H = 0.028(\pm 0.003)$, $\Lambda = 0.17(\pm 0.01)$, $\Delta = 0.26(\pm 0.02)$, $\lambda^* = 0.04(\pm 0.01)$, $s = 0.44(\pm 0.08)$

In figure 3 we also show how the predictive accuracy of the model depends on the division into calibration and validation sets. The graph shows the model's predictions for k-years ahead when the first $N - k$ years are fitted with the model. The x-axis is always the number of years k predicted. First

Variable	Prior distribution
$i(t) - m(t)$	Gauss(0,τ)
R_0	Gamma(3,1)
w_{T}	Gamma(2,2)
w_{RH}	Gamma(2,2)
Loss-of-immunity (Λ)	Gamma(1,1)
New Strain pulse (Δ)	Gamma(3,1)
New Loss-of-immunity (λ^*)	Gamma(1,1)
S_0	Uniform(0,1)
$\tau = \frac{1}{s^2}$, where s	Uniform(0,100)

Table 1: Table of prior distributions and specified values of hyper parameters.

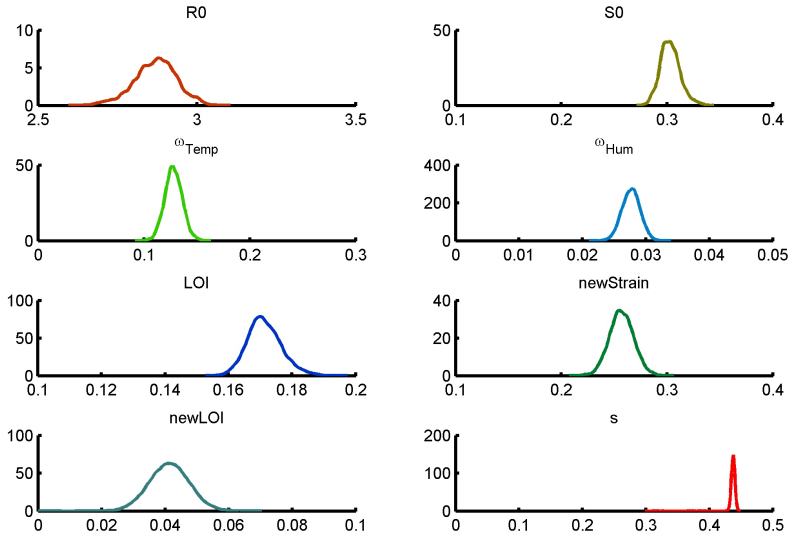

Figure 2: The density kernel plots from the trace of the MCMC procedure for sampling the posterior likelihood function. Each subpanel illustrates the estimated marginal distributions for each parameter. The confidence intervals of $\pm$ two standard deviations are readily found from this information.

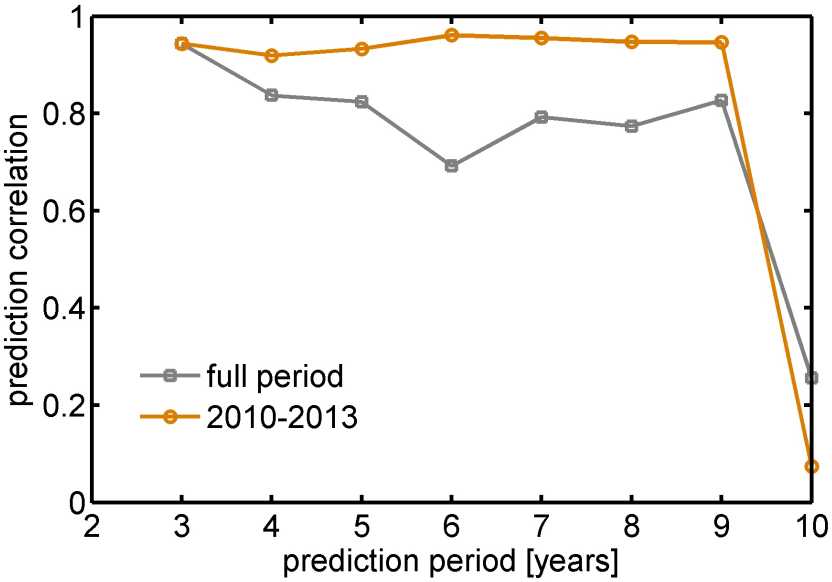

Figure 3: Model quality of fitting to $N-k$ years and prediction k years, for $N=12$. Thus $k=3$ is the reported result in Fig. 1 in the main text. $k=4$ is the then prediction quality of 2009-13, $k=5$ is 2008-13 etc. The dramatic drop-off seen for $k=10$ is due to the fact that the antigenic jump is not fitted when $N-k$=2 thus the model will fail.

we predict three years as in the paper, by fitting 12-3=9 years. Then the four years from 2009 - 2013. etc.

The brown curve shows how the original prediction of 2010-2013 behaves during that whole exercise. That is, if we fit from 2001 to $200x$ ($x<10$) then what happens to the quality of the 2010-2013 fit? It shows the same effect as in the sinusoidal v. climate driven model: adding the pulse makes the whole difference for post-antigenic jump dynamics. Note that we include 2009 in the correlation calculation for the gray curve, as opposed to Fig. 1 in the main text where we exclude this year.

3 The Data

The ILI counts and laboratory confirmed cases are in good correspondence according to Israel Center for Disease Control (ICDC)[1]. Figure 4 shows the raw data and the smoothed time series. The top panel gives the ILI daily

[1]Preceding report - influenza like illness surveillance [Internet], 2007. We will back this up with an analysis of virological surveillance data and ILI incidence counts in section 8. Available from: http://www.old.health.gov.il/Download/pages/flu_makdim2007.pdf

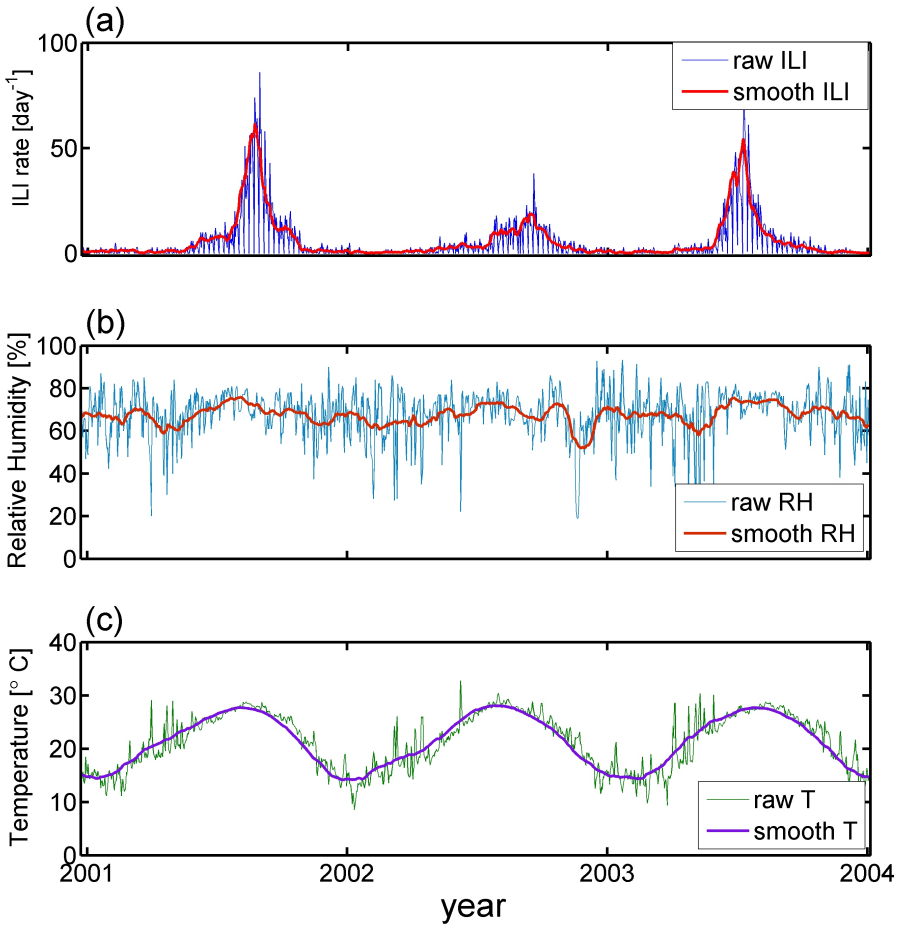

Figure 4: Examples of raw data and processed data. (a) Raw daily counts of influenza-like-illness (blue) and the smoothed data (red). The smoothing was performed with an initial linear interpolation of Saturday counts and the lower Friday counts (where opening hours are shorter too). Then a wavelet approximation is subsequently calculated. (b) Average daily relative humidity measured in and around Tel Aviv smoothed with a wavelet approximation. (c) Average daily temperature gauged in and around Tel Aviv smoothed with a wavelet approximation.

rates. The zero count 'dips' are Saturdays (Shabat) where family doctors are not taking calls. The 'spikes' are Sundays, a normal work day in Israel, whence the individuals that are still sick choose to report to their family doctor. The Saturday and Friday counts are linearly interpolated, while Sunday counts are maintained. This choice of interpolation scheme was found to be appropriate for the subsequent smoothing procedure since the weekend attack rates are preserved. A fast discrete Daubechies level 6 wavelet approximation was applied to the interpolated data. The procedure removes daily noise while maintaining fluctuations at weekly timescales (visible as wiggles in the smoothed time series in Figure 4). The climate observations were averaged for the Tel Aviv region and smoothed using a Daubechies level 6 wavelet approximation.

As we describe in Huppert et al. (2013), "[I]n our data, it was found that on average 1.5% of all Maccabi members were infected with influenza annually. However, in the US, average overall attack rates are estimated to be between 10% to 20% [3], [4], and France some 12% to 15% [5]. This, together with discussions with the Israeli Ministry of Health, motivated our setting of the reporting rate to 10% ($r = 0.1$), yielding an attack rate in Israel (ranging between 10-20%) consistent with that reported in the literature for other countries [6], [7]." Note that our premise in this work is that the coverage rate of the Health provider and the reporting rate are independent.

Regarding the population size, Israel's Central Bureau of Statistics [2] gives the population of Tel Aviv in 2008-2011 as being close to 400,000 which is what was used in the modelling.

4 The Fully Stochastic Model

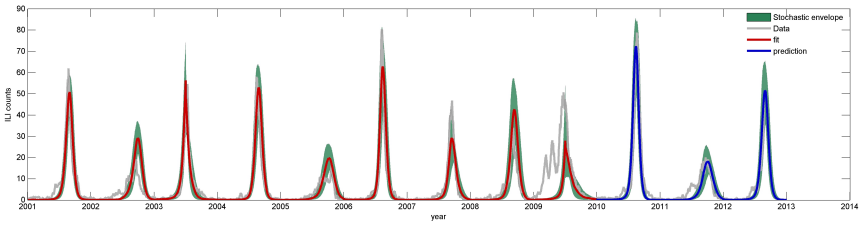

Figure 5: The ILI data (grey), the stochastic envelope as 95% prediction interval (green) and the mean (red is the fit, blue is the prediction). The quality and characterization of the fit and prediction is similar to the deterministic model in Fig. 1. in the main text.

[2] http://www.cbs.gov.il/shnaton63/st02_15.pdf

To definitively investigate the effect of demographic stochasticity we have run stochastic simulations of our model (Eq. 1) using the maximum-likelihood parameter estimates. In the stochastic simulation, the incidence at time t is given by a Poisson distribution with a mean given by the determinstic equation. This value is then used in the calculation of the incidence in the following days. The method is described in detail in [1, 8, 9]. The Poisson distribution was derived as an approximation to a generalization of the classical chain Binomial model [8]. We ran 1000 stochastic simulations of the model, calculating the mean incidence and 95% prediction interval (based on the bottom and top 2.5% incidence) for each day. The results are shown in Fig. 5 where it is clear that stochastic effects are reasonably within the conclusions we have derived with the deterministic framework.

In Fig. 6 we show the results from an exercise in validation. The stochastic model was run with the fitted parameters 100 times. Each run was then used as data for fitting and the parameter estimates were binned. The model retrieves the input parameters reasonably well with a spread that is comparable to the confidence intervals of the parameters themselves. The red stems are the input parameters (as found by fitting to the real data) and the blue histograms are the parameters estimated from fitting to the fake data created with these values.

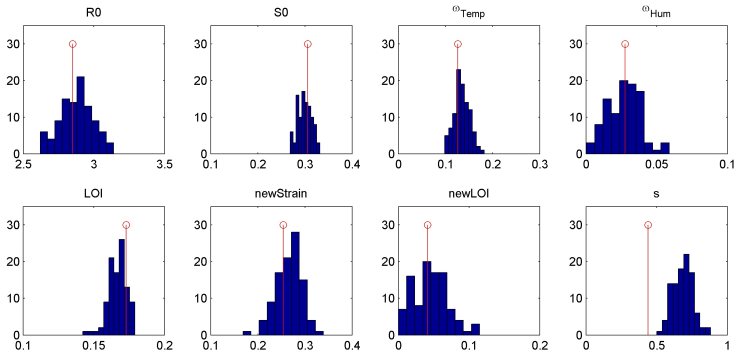

Figure 6: Histograms of parameter values found by generating trajectories of outbreaks using the values of the estimated parameters (red stems). Note that the variance $\tau = 1/s^2$ is lower due to the lower volatility of Poisson noise being compared to the reported data (see *e.g.* the outbreak in 2009/10).

5 The R_{eff} Derivation

5.1 The Basic Result

We show the derivation explaining the finding of $R_{\text{eff}} \approx 1.3$ across all flu outbreaks by analysis of the following SIRS model with population $N = 1$:

$$\begin{aligned} \dot{S} &= \Lambda R - \beta(t) SI \\ \dot{I} &= \beta(t) SI - \gamma I \\ \dot{R} &= \gamma I - \Lambda R \end{aligned}$$

Seasonal influenza outbreaks affects all age groups and thus is not crucially dependent on the low circulation rate of the population in terms of birth and deaths. In all of the analytical calculations presented here we only have seasonality as a discrete square pulse *i.e.* $\beta(t) = (1 \pm \delta)$ of total time length $4T$, where each season lasts $2T$. This choice is motivated by the finding that most outbreaks lasts approx. three months (25% of a year) and they occur in the high season, see Fig.7. We focus on build-up of susceptible population right after an outbreak where the number of infectives are low. We simplify the dynamics as:

$$\begin{aligned} \dot{S} &= \Lambda \rho \\ \dot{I} &= \beta_0 (1 \pm \delta) SI - \gamma I \end{aligned} \tag{5}$$

Here we approximate the recovered dynamics as constant $R \sim \rho$, which is reasonable for the typically low attack rates of seasonal flu.

First, following [10] we recast the variables as: $s = \frac{\beta_0}{\gamma} S$, $\omega = \log I$ and $m = \frac{\beta_0 \Lambda}{\gamma}$, which gives the following model:

$$\begin{aligned} \dot{s} &= m\rho \\ \frac{\dot{\omega}}{\gamma} &= (1 \pm \delta) s - 1 \end{aligned}$$

The infectives are then integrated during the remaining high season as:

$$\begin{aligned} \frac{\omega_1 - \omega_0}{\gamma} &= (1 + \delta) \int_{t_0}^{T} (m\rho t + c_0) dt - T^- + t_0 \\ \frac{\omega_1 - \omega_0}{\gamma} &= (1 + \delta) \frac{m\rho T^2}{2} + T[c_0(1 + \delta) - 1] - c_1 \\ \frac{\omega_1 - \omega_0}{\gamma} &= (1 + \delta) \frac{m\rho T^2}{2} + T[s_0(1 + \delta) - 1] \end{aligned}$$

The last step comes from considering that c_0 is an integration constant and $c_1 = (1 - \delta) \frac{m\rho t_0^2}{2} + t_0[c_0(1 - \delta) - 1]$. Simplifying the analysis we set $t_0 = 0$ thus $c_1 = 0$ and $c_0 = s_0$.

The build-up continues into the low season until the high season begins again:

$$\begin{aligned}\frac{\omega_2-\omega_1}{\gamma} &= (1-\delta)\int_T^{3T}(m\rho t+s_0)dt-2T\\ \frac{\omega_2-\omega_1}{\gamma} &= (1-\delta)[\frac{m\rho(3T)^2}{2}-\frac{m\rho(T)^2}{2}]+2T[s_0(1-\delta)-1]\end{aligned}$$

We can then combine these two periods and calculate the difference in infectives from the end of the outbreak to the beginning of the next:

$$\omega_2-\omega_0 = 2\frac{\delta m\rho T^2}{2}+(1-\delta)\frac{m\rho 9T^2}{2}+3T(s_0-1)-T\delta s_0$$

The null-cline for s is approximately horizontal [10, 11], which means $\omega_2-\omega_0\approx 0$. This allows for solving for s_0:

$$\begin{aligned}s_0(\delta-3) &= \delta m\rho T+(1-\delta)\frac{9}{2}m\rho T-3\\ s_0 &= m\rho T\frac{\delta+\frac{9}{2}(1-\delta)}{3-\delta}+\frac{1}{1-\delta/3}\end{aligned}$$

which in the original variables becomes:

$$S_0 = \frac{1}{R_0(1-\delta/3)}+\rho\Lambda T\frac{\delta+\frac{9}{2}(1-\delta)}{3-\delta}$$

The build-up time lasts $\tau\equiv 3T$, thus the susceptible population at the time of outbreak t_0 - which is where $\dot{S}(t_0)=0$ - is in our linear approximation simply $S_b=S_0+\rho\Lambda\tau$ according to Eq. 5. This leads to $R_{\text{eff}}\equiv R_0S_b$ being:

$$R_{\text{eff}}=\frac{1}{(1-\delta/3)}+R_0\rho\Lambda\tau\left[1-\frac{1}{3}\frac{\delta+\frac{9}{2}(1-\delta)}{\delta-3}\right] \tag{6}$$

5.2 Final approximations

We make a numerical approximation to the build-up time, τ, by simulation. Figure 7 shows the results in the interval $\delta\in[0.1,0.4]$ for R_0,S_0,α,Λ at standard values. The relationship $\tau(\delta)\approx 0.7+0.2\delta$ appear as a reasonable approximation.

We insert this approximation for τ keeping in mind that it only holds for restricted values of the other parameters. We further approximate $1/(1-\delta/3)\approx 1+\delta/3$ for δ small. Finally, Eqn. 6 becomes (with $\rho=1-|S|$ and $R_0|S|=1$):

$$R_{\text{eff}}\approx 1+\frac{\delta+(R_0-1)\Lambda(\delta+1)}{3} \tag{7}$$

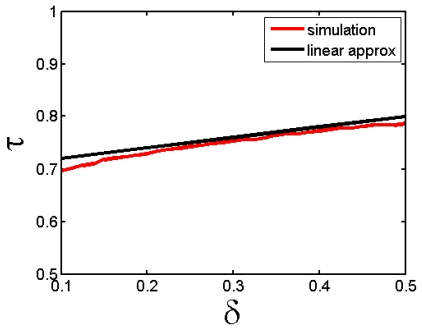

Figure 7: The simulated build up times τ for the estimated values of $\Lambda = 0.15$ and $R_0 = 3.5$ (red line). A reasonable linear approximation for $\tau(\delta)$ can then be made (black line). Note that these values of $\tau \sim \frac{3}{4}$ year are consistent with our decision to divide a year into four time periods of length T.

From this we infer that the yearly variation in the limit cycle maximal value of R_{eff} arises from changes in climate, δ, if the assumption of constant R_0 and constant loss-of-immunity rate, Λ, is valid.

For infectious diseases, the climate driver δ is typically not outside the range $[0.05, 0.3]$ see e.g. [11]. The term $1 + \delta/3$ lands values in the $[1, 1.1]$ range. The remaining term can be evaluated with reasonable values of influenza immunity loss such as $\Lambda = 0.15$ corresponding to 6-7 years of retained immunity [12, 5, 13]. The standard values for influenza basic reproduction rates are around $R_0 = 3.5$ [14]. Thus the remaining term in Eq. 7 adds another contribution of 0.14 ± 0.03 leading to R_{eff} being typically around 1.2 ± 0.1. Checking for wide R_0 variability in the range $[2, 6]$ we find that the second term gives values around 0.2 ± 0.12, which only changes the overall R_{eff} to typically around 1.25 ± 0.15.

In Fig. 8 we show that the formula in Eqn. 7 performs according to our assumptions. Except for very small seasonal amplitudes the formula is off by roughly ~ 0.02, which is an order of magnitude smaller than the variability in R_{eff} and thus negligible in the present context. Note that for the typical parameter values of influenza the simulations cannot have cycles. Without a climate driver the cycles must vanish *i.e.* $R_{\text{eff}} \to 1$ for small driver values $\delta \to 0$, which is a property not accounted for in our analytical approximation.

6 Testing the Properties of the Seasonality Driver

Figure 9 shows the effect of goodness-of-fit upon randomizing either the ILI outbreak order or the order of climate seasons. We divide the time series into

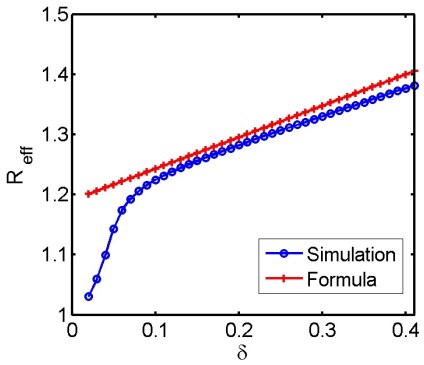

Figure 8: Numerical validation of the analytical results. We hold $\Lambda = 0.15$ and $R_0 = 3.5$ (see text), while varying $\delta \in [0.05, 0.3]$. The R_{eff} of the long term attractor at the time of outbreak is plotted against the seasonal amplitude δ (blue circles). The analytical results are plotted for comparison (red plus).

seasons stretching from June to June. Each segment, say June 2006 to June 2007, is then shifted to a random location, say June 2003 to June 2004 and the segments are exchanged. This procedure is repeated 100 times. Then we fit the period from June 2001 to June 2009 using the standard procedure outlined above. Our results imply that the order of outbreaks is the most important factor since the fit quality is always quite low $R^2 \approx 0.5 \pm 0.1$ (blue line) under randomization.

This means that the dynamics of year-to-year susceptible build-up from loss of immunity is the most important dynamical agent in the model. Following this we see that randomizing the climate seasons in a fashion similar to the ILI outbreaks leaves a mediocre fit quality at $R^2 \approx 0.7 \pm 0.1$. We then experimented with reducing the degrees of freedom by fixing parameters and scanning a range of different values while then fitting only the $S(t_0)$ and the temperature weight ω_T. The climate weight on relative humidity we fix as a ratio $\alpha = \omega_{\text{RH}}/\omega_{\text{T}} = 1/5$. The parameter search procedure shows that the model is remarkably robust to such a reduction of degrees of freedom with an average goodness of $R^2 \approx 0.83 \pm 0.05$.

7 The Physical Conditions in Tel Aviv

General evidence appears to support that climate could affect seasonal dynamics strongly in any location. There is typically a significant rise in upper respiratory infections in populations exposed to cold spells [15]. A recent study found that acute cooling of the feet causes vasoconstriction in the

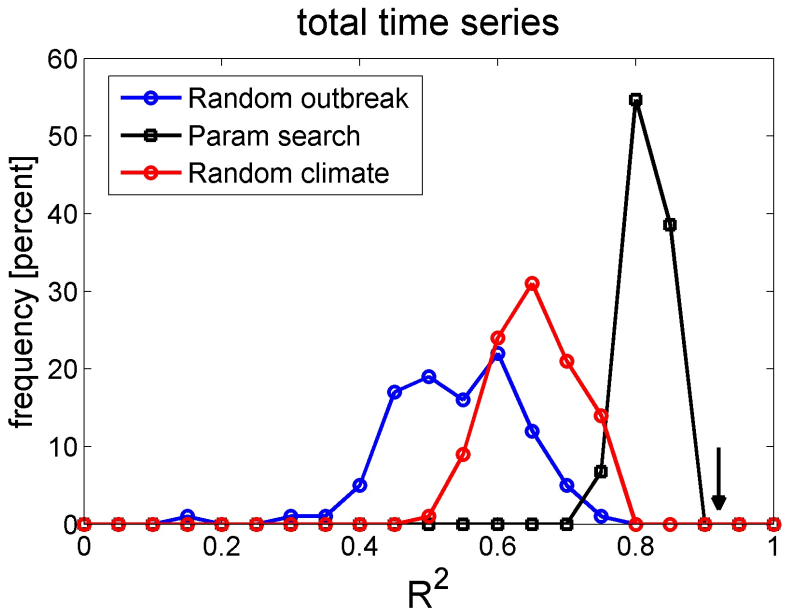

Figure 9: We perform randomizations to investigate the rank of 1) the order of outbreaks (blue), 2) the climate driver (red) and 3) scanning and fixing parameters within the ranges $R_0 \in [3, 6]$, the ratio $\omega_{\mathrm{RH}}/\omega_T \in [0.1, 0.3]$, the reporting rate $\in [0.1, 0.2]$ and the immigration rate $\in [1, 10]$ while *only* fitting S_0 and ω_T. The arrow shows the superior final full fit. We have used R^2 to emphasize the differences between the three procedures.

upper respiratory system, which potentially lowers resistance to viral infections [16]. Classic experiments have shown that mucus clearance rates of the respiratory tract are strongly reduced when inhaling cold air [17], which traps the virus particles for a longer time in the trachea allowing it a greater chance to infect the host. The influenza virus may display temperate behavior since experimental studies have shown that, independent of the host immune system, both the quantity of virus particles and the duration of peak virus shedding by infected hosts is vastly higher at 5°C than at 20°C [18]. In summary, it appears that in cold and dry conditions the virus particles survive better, are airborne for a much longer time, there are many more of them, they have easier access to cause infections and the hosts are more susceptible.

During winter a significant number of older houses in Tel Aviv and surrounding areas offer only low thermal resistance from walls, doors and window fittings. High average mid-day insolation at the Mediterranean coastline all year round discourages wide-spread use of heavy clothing during winter even though the nights are cold. Tel Aviv is a high tech business hub for workers which means that a significant part of the work force commutes there from Haifa, Jerusalem and the South on a daily basis, which is explored in a section below. At the same time Israel currently has as a low exchange of tourists and regular commuters with neighboring countries. These factors lends credibility as to why the influenza in Tel Aviv can be so uniquely modulated by local climate.

8 Relationship between ILI and influenza

In order to test the relationship between ILI and influenza, we have examined the correlation between ILI and virological surveillance data from laboratory tests of ILI cases. The laboratory tests were performed by the Central Virology Laboratory at the Tel-Hashomer hospital on sampled ILI cases collected by sentinel clinics. This data is publicly available on the WHO flunet site[3].

For each season (not including the last seasons for which we have only partial data), we have calculated the ratio of weekly ILI cases out of total ILI cases in the influenza season (weeks 40 through 10). We have calculated the Pearson correlation between this ratio and the ratio of positive laboratory tests out of the total tests performed, in weeks in which at least 10 tests were performed. The results of this procedure are presented in figure 10. For all seasons except the 2003/04 season we obtain high correlations with r between 0.69-0.95. In the 2003/04 season, during October, there were high positive percentages of influenza in the lab tests while ILI rates where still relatively low, which drove down the correlation to $r = 0.2$. This could

[3]http://www.who.int/influenza/gisrs_laboratory/flunet/en/

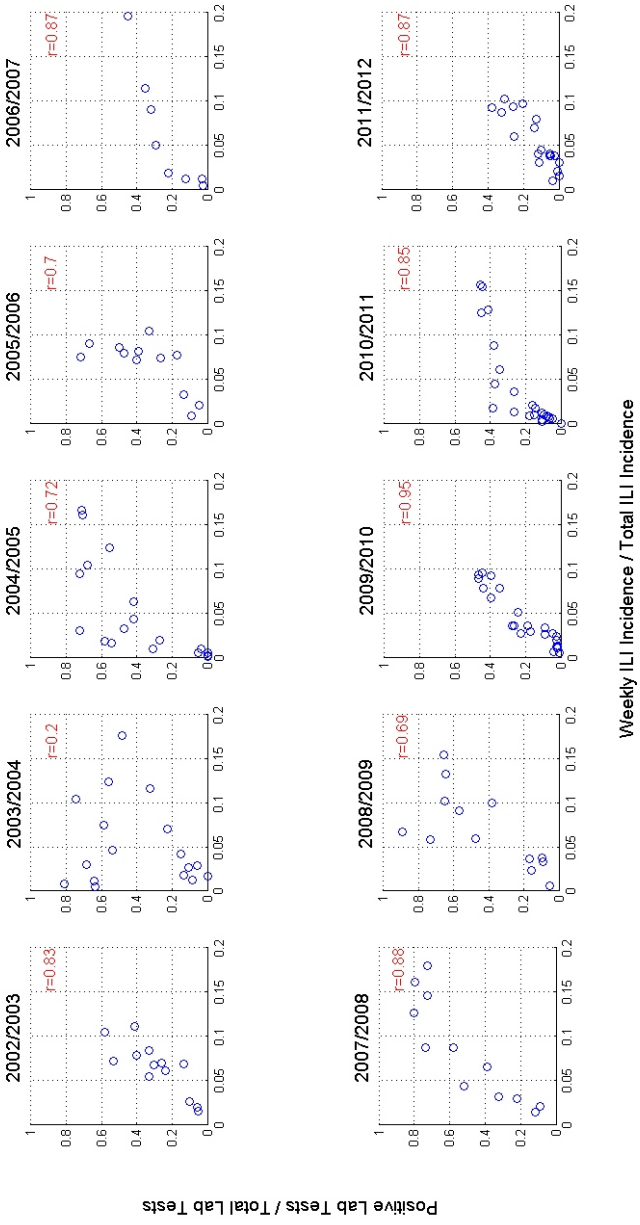
2002/2003
r=0.83
2003/2004
r=0.2
2004/2005
r=0.72
2005/2006
r=0.7
2006/2007
r=0.87
2007/2008
r=0.88
2008/2009
r=0.69
2009/2010
r=0.95
2010/2011
r=0.85
2011/2012
r=0.87
Positive Lab Tests / Total Lab Tests
Weekly ILI Incidence / Total ILI Incidence

Figure 10

be related to the fact that in this season, influenza came very early, before other causes of ILI symptoms, which is why most ILI cases were attributed to influenza, even though influenza rates were not that high yet.

The ILI time series and the lab results match the closest in the years where there were an average number of lab tests at least above 90 per week. In Figure 11 we show this property. Further, a threshold can be applied to select individual weeks of all the seasons at a lower threshold of above 75 lab tests per week. In this case only the years 2006/7, 2009/10, 2010/11. 2011/12 remain and the ILI-to-lab correlations are unchanged. This shows that we have a good correspondence between ILI and actual influenza cases for Tel Aviv since the ILI statistics is consistently collected throughout the period.

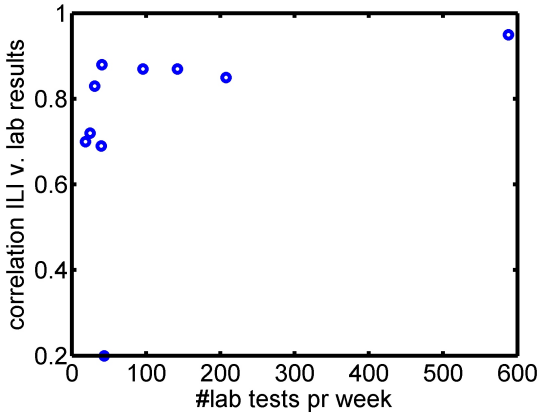

Figure 11: When the number of laboratory tests are high we see a high correspondence between ILI counts and positive influenza test results. Since the ILI registration is consistent throughout the period we conclude that the ILI time series is a valid proxy for the true influenza epidemics.

Further details of the relation between ILI and actual influenza data may also be found in [9].

9 Jerusalem influenza is driven by Tel Aviv

We have fitted the model to Jerusalem and Tel Aviv with a single R_0 and the result for Jerusalem is shown in panel (a) in the main text. We note that the outbreaks in 2005 through 2008 are not recovered exactly by the model. The overall fit quality is $r = 84\%$.

Using a coupling term whereby a fitted fraction $\psi = 30\%(5)$ of infectives of Tel Aviv is in contact with susceptibles of Jerusalem, and vice versa, the fit is improved drastically to $r = 94\%$ correlation for Jerusalem, shown in

panel (b) in main text. The model fit to Tel Aviv is almost unchanged with this coupling relatively to the independent model. This result is seemingly aligned with the fact that thousands of Jerusalem inhabitants commute every day to work in the high tech business hub of Tel Aviv and surroundings. They contract the disease in Tel Aviv and spread to the family at home in Jerusalem. The model uses this simple modification where the newly infected the follows the dynamics (with subscript T and J for Tel Aviv and Jerusalem respectively):

$$i_{\mathrm{T}}(t) = \frac{S_{\mathrm{T}}(t)}{N_{\mathrm{T}}} R_0(1+\delta_{\mathrm{T}}(t)) \sum_{\tau=1}^{d} P_\tau [i_{\mathrm{T}}(t-\tau) + \psi \cdot i_{\mathrm{J}}(t-\tau)]$$

$$i_{\mathrm{J}}(t) = \frac{S_{\mathrm{J}}(t)}{N_{\mathrm{J}}} R_0(1+\delta_{\mathrm{J}}(t)) \sum_{\tau=1}^{d} P_\tau [i_{\mathrm{J}}(t-\tau) + \psi \cdot i_{\mathrm{T}}(t-\tau)]$$

10 Numerical procedures

10.1 SIRS model fit

Assume that the ILI counts starting from January 1st 1999 are in the array `ILI`, the wavelet smoothed average daily temperature measurements are in the array `wTc` and similarly the relative humidity measurements in `wRHc`. We format arrays for the shocks given in 2003 and 2009 when new strains are introduced. The start point is `ts` for 1st June 2001 and `te` is the variable end time, with `tp` the interval length. Note that if we fit to a period shorter than 2009 we simply only fit the shock parameters to the 2003 year. The shock to the climate sensitivity lasts until the start of December (the 182 in the code below) and the shock to the loss-of-immunity starts on that same date. This date was chosen since the 2003 year peaked on that date - irregardless of the peak of following antigenic jumps.

(MATLAB script, part 1)

```
timNS = zeros(1,tp); \% shock array from new strain appearing
timNL = zeros(1,tp); \% shock from added loss-of-immunity
nsY = [3 9]; \% new strain in 2003 and 2009
for i=1:length(nsY)
    if nsY(i)*365<tp
         timNS([(nsY(i)-1)*365+1:nsY(i)*365-182])=1;
         timNL([(nsY(i)-1)*365+182:nsY(i)*365])=1;
    end
end
post = @(v) JointPostFlu(v,ILI(ts:te),...
    [-wTc(ts:te)' -wRHc(ts:te)' timNS(ts:te)' timNL(ts:te)']);
```

In order to run this script the Joint Posterior Likelihood function must be defined as below. It takes the `v` array which is the parameters to be estimated , the ILI data `Dat`, the climate (temperatures and humidity) and the arrays defining the shocks from anti-genic shocks are columns in `Clim`.

(MATLAB function)

```
function LogPr=JointPostFlu(v,Dat,Clim)

% LogPr = JointPostFlu(v,Dat,Clim)
% -Dat is incidence data of Flu
% -Clim is a matrix of climate data
% -v is vector of parameters
%   v(1) R0 is basic reproduction rate
%   v(2) S0 is initial susceptible ratio in population
%   v(3) weight on Temps (centered, negative)
%   v(4) weight on Relative humidity (centered, negative)
%   v(5) Loss-of-immunity rate
%   v(6) New Strain plateau in seasonality
%   v(7) New Strain plateau in Loss-of-immunity

% defining infectivity profile values for flu modelling
T = length(Dat);% number of days
N = 4e5*.45;    % population size TLV with a 45 % coverage
dInf = 7;       % days of infection (profile)
init = zeros(1,dInf); % zeros(1,dInf);
meanGTD = 2.285;      % previously established mean infectivity
init = zeros(1,dInf);
kGTD = 2.6;
tetaGTD = meanGTD/kGTD;
P = GetGammaGenerationTimeDistribution(dInf,kGTD,tetaGTD);

% Create simulated incidence data
rf = .1; % reporting factor
imi = 1; % immigration rate of sick persons day^-1
R0 =  v(1);
S0 =  v(2);
alpha = v(3); % temps
beta = v(4); % rel. humidity
LOI = v(5); % Loss-of-immunity per year
newStrain = v(6); % plateau for seasonality sensitivity
newLOI = v(7);    % plateau for LOI sensitivity

% define the climate array for modulating the reproduction rate
```

```
sigma = max(0,1 + alpha * Clim(:,1) + beta * Clim(:,2));

[inf0 S I R] = RunSimulation(T,N,P,init,...
    R0,S0,sigma,LOI,imi,...
    newLOI,newStrain,Clim(:,3),Clim(:,4));
inf0 = inf0*rf; % the simulated ILI time series

s = v(8); % the hyper prior on the variance estimate
tau = 1/(s*s);
pP = sum(log(normpdf(inf0,Dat',tau)));

prior = log(unifpdf(S0,0,1));
prior = prior + log(gampdf(R0,3,1));
prior = prior + log(gampdf(alpha,2,2));
prior = prior + log(gampdf(beta,2,2));
prior = prior + log(gampdf(LOI,1,1));
prior = prior + log(gampdf(newStrain,3,1));
prior = prior + log(gampdf(newLOI,1,1));
prior = prior + log(unifpdf(s,0,100));

LogPr = pP + prior;
```

The following function is required to define the infectivity profile within the Joint Posterior likelihood function.

(MATLAB function)

```
function P=GetGammaGenerationTimeDistribution(dInf,gammaA,gammaB)

pGammaCum = gamcdf(0:dInf,gammaA,gammaB);
pGamma= pGammaCum(2:dInf+1)-pGammaCum(1:dInf);
P = pGamma/sum(pGamma);}
```

The simulation that is called within the Joint Likelihood function is compiled into a MEX function using the C-code below:

(MEX function)

```
/*============================================================
 * RunSimulation.c
 *
 * Runs a simulation of an epidemic
 *
 *============================================================*/
/* $Revision: 1.0.0 $ */
```

```
#include "mex.h"
#include <stdlib.h>
#include <math.h>

/* The computational routine */

void RunSimulation(
        double* i, double* S, double* I, double* R,
        mwSize T, double N, double* P, mwSize d, double* init, mwSize initLen,
        double R0, double S0, double* sigma, double loi, double imi,
        double newloi, double nSe, double* timNS, double* timNL)
{
    mwSize t;

    for( t=initLen-1; t>=0; t-- ){ i[t] = init[t]; }
    S[0] = S0 * N;
    I[0] = i[0];
    R[0] = N - S[0] - i[0];

    for( t=1; t<d; t++ )
    {
        S[t] = S[t-1] - i[t];
        I[t] += i[t];
        R[t] = N - S[t] - I[t];
    }

    for( t=d; t<T; t++ )
    {
        double sum1=0;
        double lambda;
        mwSize tau;
        for( tau=0; tau<d; tau++ ){ sum1 += P[tau]*i[t-tau-1]; }
        lambda = S[t-1]/N*R0*(sigma[t]+nSe*timNS[t])*sum1+imi;

        i[t] = lambda;
        S[t] = S[t-1] - i[t] + (loi+newloi*timNL[t])*R[t-1]/365;
        //for( tau=0; tau<d; tau++ ){ I[t] += i[t-tau-1]; }
        I[t] = I[t-1] + i[t] - i[t-d];
        R[t] = N - S[t] - I[t];
    }
}

/* The gateway function */
```

```
void mexFunction( int nlhs, mxArray *plhs[],
                  int nrhs, const mxArray *prhs[])
{
    mwSize T;                       /* input scalar */
    double N;                       /* 1xT input matrix  */
    double *P;                      /* 1xd input matrix */
    double *init;                   /* 1xd input matrix */
    double R0;                      /* input scalar*/
    double S0;                      /* input scalar */
    double *sigma;                  /* 1xT input matrix  */
    double LOI;                     /* input scalar */
    double newStrain;               /* input scalar */
    double newLOI;                  /* input scalar */
    double imi;                     /* input scalar */
    double *timNS;                  /* 1xT input matrix  */
    double *timNL;                  /* 1xT input matrix  */
    mwSize d;
    mwSize initLen;

    double *iMatrix;               /* 1xT output matrix */
    double *SMatrix;               /* 1xT output matrix */
    double *IMatrix;               /* 1xT output matrix */
    double *RMatrix;               /* 1xT output matrix */

    /* check for proper number of arguments */
    if(nrhs!=13) {
        mexErrMsgIdAndTxt("RunSimulation:nrhs","17 inputs required.");
    }

    if(nlhs!=4) {
        mexErrMsgIdAndTxt("RunSimulation:nlhs","4 output arrays required.");
    }

    T = mxGetScalar(prhs[0]);
    N = mxGetScalar(prhs[1]);
    P = mxGetPr(prhs[2]);
    init = mxGetPr(prhs[3]);
    R0 = mxGetScalar(prhs[4]);
    S0 = mxGetScalar(prhs[5]);
    sigma = mxGetPr(prhs[6]);
    LOI = mxGetScalar(prhs[7]);
    imi = mxGetScalar(prhs[8]);
    newLOI = mxGetScalar(prhs[9]);
    newStrain = mxGetScalar(prhs[10]);
```

```
    timNS = mxGetPr(prhs[11]);
    timNL = mxGetPr(prhs[12]);
    d = mxGetN(prhs[2]);
    initLen = mxGetN(prhs[3]);
    if(initLen > d) {
        mexErrMsgIdAndTxt("RunSimulation:nlhs","init length must
        be equal or smaller than length of P.");
    }

    /* create the output matrices */
    plhs[0] = mxCreateDoubleMatrix(1,T,mxREAL);
    plhs[1] = mxCreateDoubleMatrix(1,T,mxREAL);
    plhs[2] = mxCreateDoubleMatrix(1,T,mxREAL);
    plhs[3] = mxCreateDoubleMatrix(1,T,mxREAL);

    /* get a pointer to the real data in the output matrices */
    iMatrix = mxGetPr(plhs[0]);
    SMatrix = mxGetPr(plhs[1]);
    IMatrix = mxGetPr(plhs[2]);
    RMatrix = mxGetPr(plhs[3]);

    /* call the computational routine */
     RunSimulation(iMatrix,SMatrix,IMatrix,RMatrix,
             T,N,P,d,init,initLen,R0,S0,sigma,LOI,imi,
             newLOI,newStrain,timNS,timNL);

}
```

The MEX-function above is compiled in MATLAB using the command: `mex RunSimulation.c`.

The Markov Chain Monte Carlo sampling of the likelihood function is then performed using the `slicesample` function included in the Statistics toolbox in MATLAB.

(MATLAB script, part 2)

```
initial=[3.5 .25 1.2 .5 1.2 0.1 1 0.05]; % good starting values
trace=initial;
nsamples=100000;
trace=[trace ;slicesample(trace(end,:),nsamples,'logpdf',...
    post,'width',10*ones(1,length(initial)))];
```

The array `trace` contains the sampler trace of the estimated parameters. Upon convergence of this trace an appropriately thinned part of the tail of `trace` can be averaged for extraction of values.

10.2 HHLH Motif Probability

We note that the HHL(H) motif turns out to be a generic property of our deterministic SIR model after an epochal jump. It is in fact the signature of a double-period damped oscillation, that will occur whenever there is a significant antigenic jump perturbation. Mechanistically it is reasonably clear why the model will oscillate in this fashion and relates to the dynamics of the susceptible pool (rather like a predator-prey oscillation). The biennial damped oscillation of the SIR model is also well known for measles, for example. Being generic to the model structure, the motif is not an outcome of specific fitted parameter values but is quite robust over a reasonable parameter range. (The particular form also relates to the specific climate perturbation mechanism we use). This allows us to directly estimate the probability of the motif occurring in the observed time series based on a random null hypothesis formalism, without having to rely on post-hoc or data-dredging arguments.

We first suppose that each year in the time series can be associated with coarse grained amplitudes as high, medium or low. We find it unusual that on the TWO occassions an epochal jump in virus strain appeared, namely A/Fujian in 2003 and A/H1N1 in 2009, the motif HHL(H) appeared in the ILI timeseries. [The fourth year in both motifs was either High or Medium, but not Low, which we annotate as (H). This is discussed further shortly.] We highlight this unusual occurrence in Fig. 2c in the main text where we plot the four years beginning with the A/Fujian year in 2003-4 together with the four years beginning with the A/H1N1 pandemic in 2009-10. The data seemingly collapse in the sense that the observed data in both windows sit closely on top of each other forming the same "high-high-low-(high) data-motif.

We then ask, what is the probability of two such motif patterns to occur by chance, given a time-series in which there are two (High) epochal jumps that can initiate such a motif? This can be calculated by beginning with the observed phenomenon HHL(H) and estimating the probability of this occurring under a random null hypothesis. In this scenario, the peak height in each season could be viewed as an *iid* random variable that can either be relatively high, medium or low. Beginning with H for an epochal jump year, the probability of an entire motif P[HHL(H)]= $\frac{1}{3}\frac{1}{3}\frac{2}{3} = 2/27$. The probability of two such motifs occurring if there are two antigenic jump years (and assuming the motifs do not overlap) is then P[HHL(H) $\otimes$HHL(H)]= $\frac{2}{27} \cdot \frac{2}{27} = 0.0055$.

```
function P = motifP(N,s,a)

% N is no. iterations
% s is the sequence length
```

```
% a is the alphabet size

c = 0;
for i=1:N
    S=randi(a,1,s);
    % find two jump years that allows for the motif length:
    ag1=randi(s-3,1,1); % a) with enough space at the tail, and
    ag2=randi(s-3,1,1);
    while (abs(ag1-ag2)<4) % b) with space between jumps.
        ag1=randi(s-3,1,1);
        ag2=randi(s-3,1,1);
    end
    S(ag1)=3; % Then let each jump year be epochal, and
    S(ag2)=3;
    if sum(abs(S(ag1:ag1+3)-[3 3 1 3])+abs(S(ag2:ag2+3)-[3 3 1 3]))==0
        c=c+1; % count how many times the two motifs appear.
    end
end

P = c/N;
```

References

[1] Katriel, G., Yaari, R., Huppert, A., Roll, U. & Stone, L. Modelling the initial phase of an epidemic using incidence and infection network data: 2009 H1N1 pandemic in Israel as a case study. *Journal of The Royal Society Interface* **8**, 856–867 (2011).

[2] Neal, R. Slice sampling. *The Annals of Statistics* **31**, 705–741 (2003).

[3] Cox, N. & Subbarao, K. Global epidemiology of influenza: past and present. *Annual Review of Medicine* **51**, 407–421 (2000).

[4] Glezen, W. P. *et al.* Interpandemic influenza in the Houston area, 1974–76. *New England Journal of Medicine* **298**, 587–592 (1978).

[5] Finkenstädt, B., Morton, A. & Rand, D. Modelling antigenic drift in weekly flu incidence. *Statistics in Medicine* **24**, 3447–3461 (2005).

[6] Truscott, J. *et al.* Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *Journal of The Royal Society Interface* **9**, 304–312 (2012).

[7] Chowell, G., Miller, M. & Viboud, C. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. *Epidemiology and Infection* **136**, 852–864 (2008).

[8] Katriel, G. Stochastic discrete-time age-of-infection epidemic models. *International Journal of Biomathematics* **6** (2013).

[9] Yaari, R., Katriel, G., Huppert, A., Axelsen, J. & Stone, L. Modelling seasonal influenza: the role of weather and punctuated antigenic drift. *Journal of The Royal Society Interface* **10** (2013).

[10] Stone, L., Olinky, R. & Huppert, A. Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533–536 (2007).

[11] Olinky, R., Huppert, A. & Stone, L. Seasonal dynamics and thresholds governing recurrent epidemics. *Journal of Mathematical Biology* **56**, 827–839 (2008).

[12] Levin, S., Dushoff, J. & Plotkin, J. Evolution and persistence of influenza A and other diseases. *Mathematical Biosciences* **188**, 17–28 (2004).

[13] Bansal, S., Pourbohloul, B., Hupert, N., Grenfell, B. & Meyers, L. The shifting demographic landscape of pandemic influenza. *PLoS One* **5**, e9360 (2010).

[14] Murray, J. *Mathematical biology*, vol. 2 (Springer, 2002).

[15] Eccles, R. An explanation for the seasonality of acute upper respiratory tract viral infections. *Acta Oto-laryngologica* **122**, 183–191 (2002).

[16] Johnson, C. & Eccles, R. Acute cooling of the feet and the onset of common cold symptoms. *Family Practice* **22**, 608–613 (2005).

[17] Baetjer, A. Effect of ambient temperature and vapor pressure on ciliamucus clearance rate. *Journal of applied physiology* **23**, 498–504 (1967).

[18] Lowen, A., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens* **3**, e151 (2007).