Protocol Flu prediction challenge – team Yale model

Christoph Zimmer*[1], Sequoia Leuba[1], Ted Cohen[1¶], Reza Yaesoubi [2¶],

November 7, 2016

1 Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA
2 Health Policy and Management, Yale School of Public Health, New Haven, CT, USA
* corresponding author
¶ These authors are senior authors.

1 Data sources

We use the ILInet data and the data on positive influenza tests provided by the FluView website (http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html). We multiply the weighted number of ILI cases by the percentage of positive tests from the WHO Clinical Labs report to obtain an estimate for the number of influenza cases. We also subtract this number from the number of ILI cases to obtain an estimate of ILI cases not associated to influenza. Both data sets are transformed to cases per 100000 and are used for calibration and prediction separately and added together in the end to produce a forecast.

As suggested by [1], we also use absolute humidity data from the National Land Data Assimilation System (NLDAS) project-2 dataset [2].

2 Modeling sources

We use a humidity based SIRS model developed by the Shaman group [1]. Therefore, we calculated a humidity profile for every city considered in this study. In order to obtain humidity profiles on a HHS region level, we average the humidity profiles of all cities that belong to an HHS region.

3 Calibration and prediction sources

Compartmental models

We separate the population into D mutually exclusive compartments. Let $X = (X_1, \ldots, X_D)$ denote the number of individuals in compartments $1, 2, \ldots, D$. Transitions between the groups are denoted with r and we assume that there is a total of Q transition paths. For each transition path q_{ij} denotes the number of persons leaving subgroup X_i for transition j and u_{ij} the number of persons entering subgroup X_i for transition j:

$$q_{1j}\, X_1 \;+\; q_{2j}\, X_2 \;+\; \ldots \;+\; q_{Dj}\, X_D \;\longrightarrow\; u_{1j}\, X_1 \;+\; u_{2j}\, X_2 \;+\; \ldots \;+\; u_{Dj}\, X_D, \quad \text{for } j = 1, \ldots, r$$

The so called stoichiometric matrix Γ is a $D \times r$ dimensional matrix. Its entries $s_{ij} = u_{ij} - q_{ij}$ describe the net effect of transition j to subgroup X_i. In terms of ODEs the system would read as

$$\frac{d}{dt}x(t;\theta,x_0) \;=\; \Gamma\ \Lambda(x(t;\theta,x_0),\theta),\ \text{with } x(0,\theta,x_0) = x_0 \tag{1}$$

with a rate law $\Lambda = (\Lambda_1, \ldots, \Lambda_r)^T$ describing the speed of the transitions, an initial value x_0 and a parameter vector θ.

Stochastic modeling is important in systems with small numbers of people in at least one of the subgroups, where stochastic fluctuations may have a major influence on the disease dynamics. Stochastic modeling can model each transition of a person from one subgroup to another explicitly, by choosing the transition and the time of the transition stochsatically. Both order of transitions and waiting times are stochastic quantities depending on the system's state and the rate laws. The master equation (ME, equation 2) describes the time evolution of the probability of the system to be in a state ν:

$$\begin{aligned} \frac{d}{dt}P_\theta(\nu,t|\nu_0,t_0) \;&=\; \sum_{r=1}^{Q}\Big(\tilde{\Lambda}_r(\nu - s_r,\theta)P_\theta(\nu - s_r,t|x_0,t_0) - \tilde{\Lambda}_r(\nu,\theta)P_\theta(\nu,t|\nu_0,t_0)\Big) \\ P_\theta(\nu,t_0|\nu_0,t_0) &= \begin{cases} 1 & \nu = \nu_0 \\ 0 & \text{else} \end{cases} \end{aligned} \tag{2}$$

with a vector $s_j = (s_{1r}, \ldots, s_{Dr})^T$ and with a propensity $\tilde{\Lambda}$ that can be calculated from the rate law Λ and describes the speed of the reactions in terms of particle numbers.

The Gillespie algorithm [3] is the method of choice to simulate stochastic time courses. It is an iterative algorithm simulating reaction event after reaction event using functions of random numbers to determine both time step and reaction. The resulting time course is then a discrete state continuous time Markov jump process.

Calibration

Observations $y_1, \ldots, y_n$ are obtained at time points $t_1, \ldots, t_n$. These might be for example weekly diagnoses of new cases. However, there is no need for equidistant observations. We use an iterative procedure to update our knowledge with each new observation. Let us assume that we can summarize our prior knowledge on the parameters before the first observation in a prior distribution $\pi_0(\theta)$. We use each new observation y_i to update our knowledge on the parameter θ by multiplying our prior with the probability to observe y_i:

$$\pi_i(\theta|y_1,\ldots,y_{i-1},y_i) \;=\; \pi_{i-1}(\theta|y_{i-1},\ldots,y_1)\mathcal{P}(y_i|y_1,\ldots,y_{i-1};\theta) \tag{3}$$

Note that the posterior at time t_{i-1}, π_{i-1}, is as well the prior at time t_i. For $i = 1$ we set $\pi_0(\theta|y_0) = \pi_0(\theta)$ as we do not have an observation at time t_0. The two following subsection will explain a) how to set up a suitable approximation for $\mathcal{P}$ and b) how to propagate a sample of the distribution π_i through time.

MSS method

We will use the multiple shooting for stochastic system's method to approximate $\mathcal{P}$ in a way that is fast enough to be computational feasible but still accurate enough to allow for reliable calibration and prediction. The MSS method has been developed in a systems biology context [4–6] and already successfully applied to calibration and prediction in epidemics [7]. This subsection will briefly summarize it, for details we refer the reader to the original publications.

Let $\Pi(\cdot|Y_i)$ denote the belief state at time t_i given the accumulated observations Y_i. Now by conditioning on the epidemic state at time t_{i-1}, i.e. ν_{i-1}, the probability function $\mathcal{P}(\cdot|y_1, y_2, \ldots, y_{i-1}; \theta)$ in Eq. (3) can be calculated as:

$$\mathcal{P}(y_i|y_1, y_2, \ldots, y_{i-1}; \theta) = \sum_{\nu_{i-1} \in \Omega_{i-1}} P(y_i|\nu_{i-1}; \theta)\ \Pi(\nu_{i-1}|y_1, y_2, \ldots, y_{i-1}; \theta), \tag{4}$$

where Ω_{i-1} is the support of the belief state at time t_{i-1}. By conditioning on the state of the epidemic at time t_i, the probability function $\mathcal{P}(\cdot|y_1, y_2, \ldots, y_{i-1}; \theta)$ in Eq. (4) can be calculated as:

$$\begin{aligned}&\mathcal{P}(y_i|y_1, y_2, \ldots, y_{i-1}; \theta)\\ &\quad= \sum_{\nu_i \in \Omega_i} \sum_{\nu_{i-1} \in \Omega_{i-1}} P(y_i|\nu_i, \nu_{i-1}; \theta) p(\nu_i|\nu_{i-1}; \theta)\ \Pi(\nu_{i-1}|y_1, y_2, \ldots, y_{i-1}; \theta).\end{aligned} \tag{5}$$

Calculating the probability function (5) can be computationally difficult. First, it requires calculating or approximating the transition probability $p(\nu_i|\nu_{i-1}; \theta)$ for each pair $(\nu_i, \nu_{i-1}) \in \Omega_i \times \Omega_{i-1}$, and second, it involves enumeration over the set $\Omega_i \times \Omega_{i-1}$, which can be prohibitively large even for simple epidemic models.

We will first describe a way to approximate p such that it is still accurate enough for calibration and prediction but also computationally fast enough to be applied to realistic size models. After that we will address the belief state Π.

Approximating state transition probabilities

Finding the exact state transition probability function $p(\cdot)$ can be difficult, and in many cases impossible, as state spaces in epidemic models can be quite large or unbounded. To overcome this problem, we employ a linear noise approximation (LNA) method to approximate the probability distribution of the new epidemic state ν_i given the previous state ν_{i-1}, i.e. $p(\nu_i|\nu_{i-1}; \theta)$. The LNA has been previously used to estimate parameters of stochastic biochemical reaction models [5, 6]. Here we extend Zimmer and Sahleâs method [6] to calibrate stochastic epidemic models where the true epidemic state is only partially observable.

To approximate the probability distribution of ν_i given the state ν_{i-1} , the LNA method uses an ordinary differential equations (ODE) model to approximate the *expected* behavior of the epidemic over the period $[t_{i-1}, t_i]$ and to identify a co-variance matrix to characterize the uncertainty around the epidemic behavior over this interval. We use the following notation to denote the ODE epidemic model used by the LNA method:

$$\begin{aligned}\frac{d}{dt}x(t, x_0; \theta) =& \Gamma\Lambda(x(t, x_0; \theta), \theta),\\ x(0, x_0; \theta) =& x_0.\end{aligned} \tag{6}$$

In the ODE system (6), the vector $x(t, x_0; \theta)$ is the epidemic state of the ODE model at time t given the initial state x_0, the vector $\Lambda(x(t, x_0; \theta), \theta)$ denotes the instantaneous changes in the

epidemic when at state $x(t, x_0; \theta)$, and the matrix Γ describes how the instantaneous changes at time t impact the epidemic state at time $t + \Delta t$ (see subsequent sections for an example).

The LNA assumes that the probability distribution of $\nu_i|\nu_{i-1}$ can be properly approximated by a normal distribution $\mathcal{N}(\mu_i, \text{cov}_i)$. The mean vector μ_i is the solution of ODE system (6) with ν_{i-1} as the initial condition (i.e. $\mu_i = x(t_i - t_{i-1}, \nu_{i-1}; \theta)$) and the variance matrix $\text{cov}_i = \Sigma(t_i - t_{i-1}, \nu_{i-1}; \theta)$ is the solution of the following ODE systems [8, 9]:

$$\begin{aligned} \frac{d}{dt}\Sigma(t, \nu_{i-1}; \theta) &= J(x, \theta)\Sigma(t, \nu_{i-1}; \theta) + \Sigma(t, \nu_{i-1}; \theta)J(x, \theta)^T + D(x; \theta) \\ \Sigma(0, \nu_{i-1}; \theta) &= 0_{K\times K}. \end{aligned} \tag{7}$$

In the ODE system (7), $J(x, \theta) = \Gamma \frac{d}{dx}\Lambda(x, \theta)$ and D is a $K \times K$ matrix with the (i, j) entity equal to $\sum_{k=1}^{K} \Gamma_{jk}\Gamma_{jk}\Lambda(x, \theta)$, where K is the number of compartments in the epidemic model.

An important question remains about how well this proposed LNA method approximates the probability distribution of epidemic states. Relying on an extensive numerical analysis, we will demonstrate in the Results section that for the epidemic scenarios considered, our method yields accurate parameter estimations and reliable predictions. We also note that our method does not rely on a single LNA model to approximate the entire epidemic trajectory. For each observation period $[t_{i-1}, t_i]$, it generates a new LNA model to approximate the epidemic behavior only over this particular period.

Updating belief states

It remains to address the belief state Π. We will first describe one possibility according to previous work on the MSS method. This idea is more general is it can be used with general observation probabilities P. The second is specificically designed for gaussian observation noise.

One way to simplify the computational complexity of Eq. (5) is to represent the belief state $\Pi(\cdot|Y_i)$ as a step function that takes 1 for the most probable state (denoted by $\hat{\nu}$) and 0 elsewhere. This allows us to approximate the function $\mathcal{P}(\cdot)$ in Eq. (5) with:

$$\tilde{\mathcal{P}}(y_i|y_1, y_2, \ldots, y_{i-1}; \theta) = \sum_{\nu_i \in \Omega_i} P(y_i|\nu_i, \hat{\nu}_{i-1}; \theta)p(\nu_i|\hat{\nu}_{i-1}; \theta), \tag{8}$$

where $\hat{\nu}_{i-1}$ represents the most likely epidemic state given observations $Y_{i-1} = (y_1, y_2, \ldots, y_{i-1})$. By the definition of epidemic states, the transition from state ν_{i-1} to ν_i generates a unique set of observations, and it it trivial to find whether the state transition $\hat{\nu}_{i-1}$ to ν_i can generate the observation y_i (see design of performance analysis for an illustrative example). Therefore, for a given observation y_i, the probability $P(y_i|\nu_i, \hat{\nu}_{i-1}; \theta)$ in Eq. (8) is equal to 1 if the transition from state $\hat{\nu}_{i-1}$ to ν_i results in observing y_i, and is zero otherwise. To calculate $\tilde{\mathcal{P}}(y_i|\hat{\nu}_{i-1}; \theta)$ in Eq. (8), it only remains to identify the state transition probability function $p(\nu_i|\hat{\nu}_{i-1}; \theta)$ and a state estimation scheme.

We now describe how to update our belief about the true state of the epidemic, denoted by $\Pi(\cdot)$, once new observations y_i are obtained. We first note that $\Pi(\cdot)$ is defined to return 1 for the most likely state $\hat{\nu}$, and zero elsewhere. Therefore, given the state $\hat{\nu}_{i-1}$ at time t_{i-1}, the probability of observing y_i during the interval $[t_{i-1}, t_i]$ is equal to $P(y_i|\nu_i, \hat{\nu}_{i-1}; \theta)p(\nu_i|\hat{\nu}_{i-1}; \theta)$ (see the discussion

prior to Eq. (5)).

Now, the most probable state at time t_i, $\hat{\nu}_i$, is the one that leads to the highest probability of observing y_i:

$$\hat{\nu}_i = \underset{\nu_i \in \Omega_i}{\arg\max}\, P(y_i|\nu_i, \hat{\nu}_{i-1}; \theta) p(\nu_i|\hat{\nu}_{i-1}; \theta). \tag{9}$$

Re-sampling to avoid filter degeneracy

The above described methodology to evaluate the probability to observe a sequence $y_1, \ldots, y_n$ can be very flexibly combined with various optimization techniques such as global optimization, gradient based or Bayesian approaches. Here, we choose a Bayesian approach. A posterior distribution π_i is calculated based on a prior π_{i-1} and its update by the probability $P(y_i|y_1, \ldots, y_{i-1}; \theta)$ as in equation (3). Filtering techniques [1, 10] update a sample from the prior distribution with the probability of the current observation to calculate a posterior distribution.

Therefore, we initially sample M parameters $\theta_1, \ldots, \theta_M$ from a prior parameter distribution π_0 and M initial states $\nu_{0,1}, \ldots, \nu_{0,M}$ from a prior state distribution $\Pi(\nu_0)$. We consider the initial states as an additional dimension of the parameter and denote $\vartheta = (\theta, \nu_0)$. For each of the vectors $\vartheta_m, m = 1, \ldots, M$ we calculate the probabilities $\pi_i(\vartheta_m)$ recursively. This results in a weighted sample of parameters $\vartheta_1, \ldots, \vartheta_M$ with weights $w_1, \ldots, w_M$.

One potential drawback of this procedure is so called filter degeneracy. We use a finite set of parameters $\vartheta_1, \ldots, \vartheta_M$ as a sample of our prior distribution and weight it with the probability of each new upcoming observation. After some observations and weighting it might happen that only very few parameters ϑ_m have some weight and all the others have weights that are several orders of magnitude lower. In the extreme scenario, only one parameter vector ϑ is left and all others have negligible weights. This is a serious issue for calculating confidence intervals or performing reliable predictions as it means that our posterior is basically a point distribution.

To deal with it, one needs two things: 1) a mechanism to detect it and 2) a mechanism to resolve it.

Detection of filter degeneracy is relatively simple. One needs to calculate which fraction of parameter vectors substantially contribute weights to the overall sample. One way to do this is described in [11]: As a first step, the weights are normalized:

$$\tilde{w_m} = \frac{w_m}{\sum_{j=1}^{M} w_j}$$

Then, the so called effective sample size N_{eff} is estimated as a measure of (non-)degeneracy (comp geo paper):

$$\hat{N}_{\text{eff}} = \frac{1}{\sum_{m=1}^{M} w_m^2} \tag{10}$$

In the beginning all particles have equal weight and, hence, N_{eff} is M. In the other extreme, if all weight is on one particle, N_{eff} is 1. Therefore, one way of detection of degeneracy is to determine a threshold value N_C below which one takes measures to re-sample the particles. The calculation of N_{eff} it not time-consuming and, hence, one can easily perform this check in every iteration.

It is unfortunately a bit more difficult, how to **resolve the filter degeneracy**. We will present one approach which was successful for us but note that IMIS [12] presents an alternative. As soon as filter degeneracy is detected, re-sample L parameters $\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_L$ from $\vartheta_1, \ldots, \vartheta_M$ according to current weights $w_1, \ldots, w_M$. Use these sample and a kernel density estimation to calculate

$$\hat{f}(\vartheta) = \frac{1}{Mb} \sum_{l=1}^{L} K\left(\frac{\vartheta - \tilde{\vartheta}_l}{b}\right) \tag{11}$$

with a bandwidth b and a kernel K.

Next, M new points are sampled from this density $\hat{f}$ and the new weights are calculated with

$$w_m = \prod_{k=1}^{i} \mathcal{P}(y_i | y_1, \ldots, y_{i-1}, \theta_m) \tag{12}$$

with i being the current iteration and the m-th initial state estimate $\nu_{0,m}$ determined by the state component of ϑ_m.

1. Initialization
   (a) Choose an initial prior probability function $\pi_0(\theta)$.
   (b) Choose an initial set of parameters $\{\theta_1, \ldots, \theta_l\}$.
2. Calibration: For each observation y_i, $i \in \{1, 2, \ldots\}$,
   (a) Calculate the probability $\tilde{\mathcal{P}}(y_i | Y_{i-1}; \theta_m)$ for every $m = 1, \ldots, M$ (using Eq. (8) and the method we described).
   (b) Update the parameter posterior distribution:
   $\pi_i(\theta | Y_i) \leftarrow \tilde{\mathcal{P}}(y_i | Y_{i-1}; \theta_m) \pi_{i-1}(\theta | Y_{i-1})$ for every $m = 1, \ldots, M$.
   (c) Update the belief state $\Pi(\nu_i | Y_i; \theta_m)$ for every $m = 1, \ldots, M$ (through solving optimization problem (9)).
   (d) Check whether degeneracy occurs by comparing the effective sample size N_{eff} (10) with a threshold N_C.
   (e) If step (d) detects filter degeneracy: use a density estimate $\hat{f}$ (11) to re-sample new parameters.

Figure 1: An algorithm for real-time calibration of stochastic compartmental epidemic models including mechanisms against filter degeneracy

Prediction

Predictions are done based on the posterior distribution of the parameters and states and with the help of the simulation model. Let us denote our prediction target with Z where Z can be a

very general target (e.g. next weeks cases or peak timing). We are interested in calculating the probability density $P(Z = z)$. This can be done by:

$$P(Z|Y_i) = \int_{\theta\in\Theta} \int_{\nu_i\in\Omega_i} P(Z|\nu_i;\theta)\ \Pi(\nu_i|Y_i;\theta)\ \pi_i(\theta|Y_i)\ d\nu_i\ d\theta. \tag{13}$$

where $P(Z|\nu_i;\theta)$ is the probability to observe Z given ν_i and θ. This probability can be calculated with the help of the simulation model.

Re-sample $\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_{\bar{L}}$ from the current weighted sample $\vartheta_1, \ldots, \vartheta_L$ according to its weights $w_1, \ldots, w_L$. Carry out $\bar{L}$ simulations initialized with the state estimate $\hat{\nu}_i$ that corresponds to the parameter θ_l until the time of the target Z.

4 Specifications

- The humidity based SIRS model has four parameters, L, D, $R0max$, $R0min$ and two initial states: S_0 and I_0. We use 1000 parameter to cover this 6 dimensional space, initially uniformly distributed on $[365, 3650] \times [1.5, 7] \times [1.3, 4] \times [0.8, 1.2]$) and $[50000, 70000] \times [5, 100]$ for the Flu data and $[50, 1000] \times [1.5, 7] \times [1.3, 4] \times [0.8, 1.2]$) and $[70000, 90000] \times [500, 5000]$ for ILI minus FluView
- We use an additional Gaussian observation noise of standard deviation 10.
- We use 100 draws from the posterior to run the simulation model to obtain predictions.
- Density estimation for re-sampling: we use Mathematica's Gaussian mixture kernel function with max 5 mixture kernel (for run time reasons) and a bandwidth of 100 for L, 0.5 for D, 0.5 for $R0max$, 0.025 for $R0min$, 1000 for S_0 and the median of the re-samples for I_0.
- Density estimation for target posteriors: We use Mathematica's Smooth Kernel Distribution density estimation function.

Historical priors are calculated for the target Season Onset", "Season Peak Week" and "Season Peak Percentage" by taking the data from years 1997/1998 to 2015/2016 without the atypical 2009/ 2010 season, the Mathematica function Smooth Kernel Distribution is used here as well for density estimation. Additional tuning is performed for these targets as follows:

- Season Onset: if 99% quantile of posterior smaller than minimum of historical prior OR 1% quantile of posterior greater than maximum of historical prior, use historical prior, else, truncate posterior with minimum and maximum of historical priors.
- Season Peak Week: if epidemic week no later than week 6, then, historical prior otherwise target posterior.
- Season Peak Percentage: if 99% quantile of posterior smaller than minimum of historical prior OR 1% quantile of posterior greater than maximum of historical prior, use historical prior, else, truncate posterior with minimum and maximum of historical priors.

References

[1] W. Yang, A. Karspeck, and J. Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLOS Computational Biology*, 10:e1003583, 2014.

[2] Y Xia and et al. Nldas primary forcing data l4 hourly 0.125 x 0.125 degree v002. *, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC)*, Accessed August 8th 2016.

[3] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22 (4):403–434, 1976.

[4] C. Zimmer and S. Sahle. Parameter estimation for stochastic models of biochemical reactions. *Journal of Computer Science & Systems Biology*, 6:011–021, 2012.

[5] C. Zimmer and S. Sahle. Deterministic inference for stochastic systems using multiple shooting and a linear noise approximation for the transition probabilities. *IET Systems Biology*, 9:181 – 192, 2015.

[6] C. Zimmer. Reconstructing the hidden states in time course data of stochastic models. *Mathematical BioSciences*, 269:117 – 129, 2015.

[7] C. Zimmer, R. Yaesoubi, and T. Cohen. Estimating the effective reproductive number for a novel viral pathogen using a stochastic compartmental model. *Epidemics 5, Clearwater Beach*, December 2015.

[8] P. Thomas, H. Matuschek, and R. Grima. Intrinsic noise analyzer: A software package for the exploration of stochastic biochemical kinetics using the system size expansion. *Plos ONE*, 7:e38518, 2012.

[9] N. G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2007.

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

[11] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 50:174–188, 2002.

[12] A. E. Raftery and L. Bao. Estimating and projecting trends in hiv/aids generalized epidemics using incremental mixture importance sampling. *Biometrics*, 66:1162–1173, 2010.