Protocol Flu prediction challenge – team Yale model 1

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

August 23, 2017

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 provided by the FluView website (http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html). In addition we use the ILINearby website data [1]. As suggested by [2], we also use absolute humidity data from the National Land Data Assimilation System (NLDAS) project-2 dataset [3].

2 Modeling sources

We use a humidity based SIRS model developed by the Shaman group [2]. 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

We use our Multiple Shooting for Stochastic Systems (MSS) framework as in [4] (MSSa version) with the following add-ons:

3.1 Weighting between historical experience and model based predictions

Motivated by the work of [5], we decided to use a weighting between historical experience and model based predictions. For this year's participation, we start with a simple weighting scheme as follows: We use the available CDC ILI data [6] to set up probability distribution on the long-term forecasting targets based on historical knowledge. For the onset we use years 2007 onwards (that's where baseline data is easily available) without pandemic, for peak time and intensity from 2005 onwards without pandemic.
We calculate weights for each HHS region, forecasting target and forecasting week, based on performance of history based forecast and model based forecasts of the years 2013, 2015, 2016. Denoting

the historical based forecasting distribution as P_h and the model based forecasting distribution as P_m, we determine the weight w such as to maximize the logscore of the weighted forecasting distribution $(1-w)P_h + wP_m$.

3.2 How to combine ILINearby data and CDC ILI

We use CDC ILI for the past epidemic weeks and add one data point from ILINearby at the end of the ILI time series data for the epidemic week for which ILINearby is already available but CDC ILI is not yet.

In order to account for uncertainty in ILINearby nowcast, and also for potential corrections in CDC ILI after its publication, we do the following: When combining the time series data of the past and our model based forward simulation into the future (e.g. in order to forecast onset or peak), we add noise to the last three time series data points, namely the third and second last data point (which is from CDC ILI) and the last data point (which is from ILINearby). The three variances for these three time points come from the ILINearby data source for these three weeks.

3.3 Specifications

- We use 1000 samples to describe the parameter distribution and 100 draws from the posterior to run the simulation model to obtain predictions.
- We use an additional Gaussian observation noise of standard deviation 10.
- 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.
- From EW 8 onwards: truncate peakintensity distribution at 90% of so far observed peak. Reason: The Kernel density estimate might in some cases place weight below and we do have observed higher data values. 90% because CDC ILI might later on be corrected so we want to be conservative.
- Similar for peak time with using time point when 90% of so far observed peak occurred.
- 1w-4w: truncation at 13100 as historical experience shows that unlikely to be higher.

References

[1] Delphi research group at Carnegie Mellon University. *http://delphi.midas.cs.cmu.edu/nowcast/.*

[2] 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.

[3] 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.

[4] C. Zimmer, S. Leuba, R. Yaesoubi, and T. Cohen. Accurate uncertainty quantification in epidemic parameter estimation and prediction using stochastic compartmental models. *submitted to Statistical Methods in Medical Research*, submitted July 2017.

[5] E. L. Ray and N. G. Reich. Prediction of infectious disease epidemics via weighted density ensembles. *arXiv*, 1v63901.3071:viXra, 2017.

[6] Centers for Disease Control and Prevention. Flu activity and surveillance. *https://www.cdc.gov/flu/weekly/fluactivitysurv.htm*, Accessed May 31st, 2017.