**Team LANL: Forecasting Methodology**

Dave Osthus and Sara Del Valle

11-7-16

LA-UR-16-28596

We fit a probabilistic state-space model motivated by the deterministic Susceptible-Infectious-Recovered (SIR) model and apply a Bayesian modeling approach. We specify an informative prior on the parameters of the model based on model fits to historical flu seasons. The informative prior specification allows us to produce forecasts consistent with historical flu seasons in the absence of data for the current flu season. As data become available, the model is updated (i.e., posterior parameter distributions are modified and constrained) and forecasts are adjusted accordingly. Forecasts are computed as functions of realizations from the posterior predictive distribution of the Bayesian model.

We have found that a probabilistic state-space model with an SIR-motivated process equation plus a random error component (i.e., noise) does not adequately account for the trends of the weighted Influenza-Like-Illness (ILI) data. For this reason, our model includes a discrepancy component to account for the *systematic* deviations between the SIR-motivated curve and the data. The discrepancy model is a first-order, reverse random walk that provides flexibility and smoothness in time. The reverse random walk encourages forecasted flu trajectories to revert to levels that are historically consistent by the end of the flu season (roughly the end of May). We found that without this specification, our model’s forecasts were unrealistically high at the end of the flu season. This was especially true for late peaking flu seasons.

Two data sources are used: weighted ILI and English language Wikipedia access logs to influenza-relevant pages. The weighted ILI is used for forecasting the flu nationally and regionally (the regions are based on the Health and Human Services Regional Health administrations). Wikipedia is used to augment weighted ILI for our national forecast. As Wikipedia access logs are available in near real-time and weighted ILI is released at a 1 to 2 week lag, Wikipedia provides a unique source of information for our model. Wikipedia is not used in our regional forecasts because geographic information is not attached to Wikipedia access logs.