Conditioning Multi-model Ensembles for Disease Forecasting

J. Ray and S. Lefantzi

Sandia National Laboratories, Livermore, CA

OBJECTIVE

Develop methods that use observational data to reconcile conflicting predictions generated by an ensemble of disease models

- Improve forecasting of influenza in the San Francisco Bay Area (SFBA)
- Use ILI+ derived from Google Flu Trends (GFT) available at SFBA cities; also meteorological data (from National Land Data Assimilation System)
- Use an ensemble of 5 models with very different underlying modeling assumptions & techniques

BACKGROUND

- There are a large variety of disease models; many are represented in the Biosurveillance Ecosystem (BSVE)
- Diverse data streams, and not just public health data, can be used to track disease outbreaks
  - E.g., ones derived from logs of web searches (Google Flu Trends), Twitter posts etc.
- Data assimilation systems e.g., ensemble Kalman filters and Bayesian techniques have been used to calibrate these models to observational data
- Despite being calibrated to data streams, these model ensembles provide conflicting forecasts for a given outbreak
- The reasons are:
  - Shortcomings / quality of the calibrating data
  - Relevance of the modeled epidemiological processes in the given outbreak
  - Redundant data streams and/or models that provide no new information to the ensemble

So, how do we improve the ensemble?

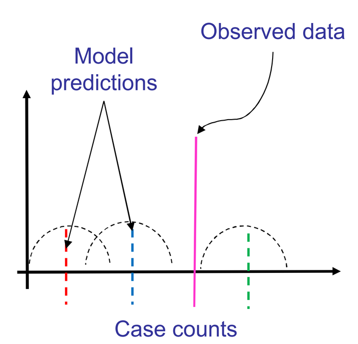
- Identify and remove the data streams & models that are redundant
- Up-weigh the more predictive models (all models are not equal, in a given outbreak)
  - Obviously, we will need "ground truth" i.e., observational data
- Given data, also provide "error bounds" for each model prediction

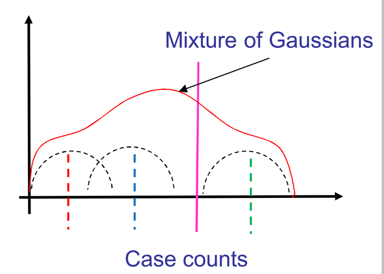
TECHNICAL APPROACH

- The key to reconciling *N* conflicting model predictions $f_n(t)$ for time t is to consider the ensemble prediction to be a weighted mean of the model predictions
- $Y^{(ens)}(t) = \sum_n \alpha_n f_n(t)$, where α_n are weights that sum to 1
- α_n are estimated so that $Y^{(ens)}(t)$ is similar to observations $Y^{(obs)}(t)$
- Or we can formulate the problem to keep all models but severely down-weight the ones that do not contribute to predictive skill e.g., via Bayesian Model Averaging
  - Requires some observational data
- The formulation of the estimation problem can be designed to remove redundant/correlated models and data streams e.g. via shrinkage
  - Useful if we lack observational data

Bayesian Model Averaging (BMA)

- Assume that model predictions $f_n(t)$ are actually means of a Gaussian distribution $G(f_n(t), \sigma^2)$; variance σ^2 is unknown
- $Y^{(obs)}(t) \sim \sum_n \alpha_n G(f_n(t), \sigma^2)$ is a draw from a mixture of Gaussians
- We seek α_n that maximizes the likelihood of the draws $Y^{(obs)}(t)$, $t = 1 \ldots T$; solved via Expectation Maximization





Shrinkage

- Fit the ensemble to data (estimate α_n) by shrinkage
- $\min_{\alpha_n} \left\| Y^{(obs)}(t) - \sum_n \alpha_n f_n(t) \right\|_2^2 + \lambda \|\alpha_n\|_1$
- λ ensures that many α_n get driven to zero; optimal λ determined by cross-validation. Use LASSO to do so

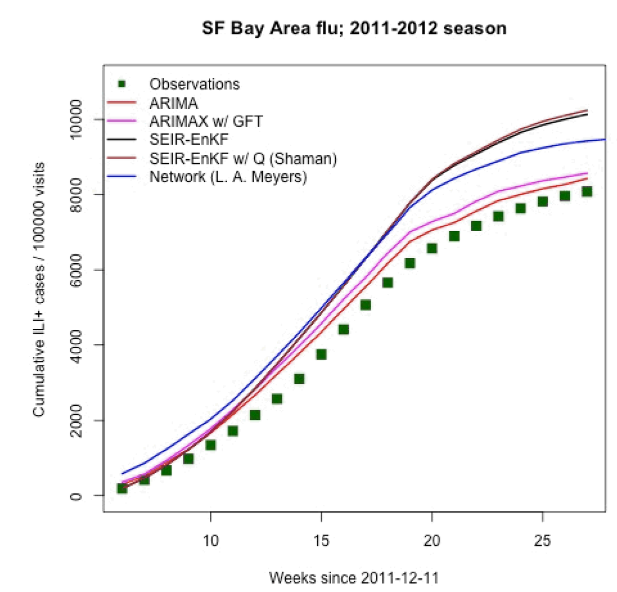
Test case

- Improve forecasting of influenza in the SFBA using an ensemble of 5 models
- Data: ILI+, derived from Google Flu Trends multiplied by % of ILI cases testing positive for flu. Use 2010-2013 data
- Randomly split data equally into a set for training individual models and another for model averaging

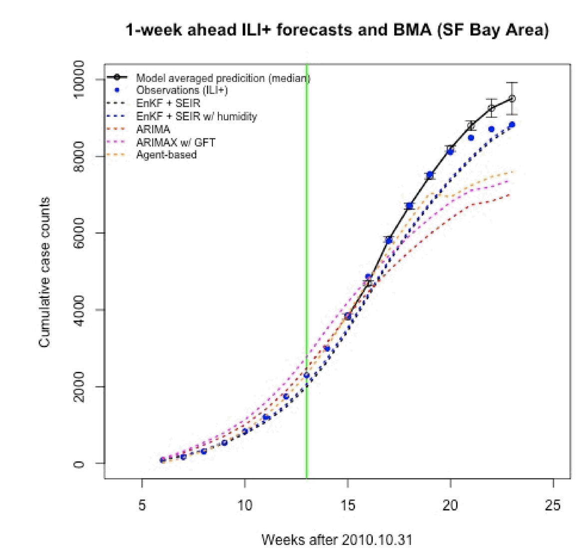
THE ENSEMBLE

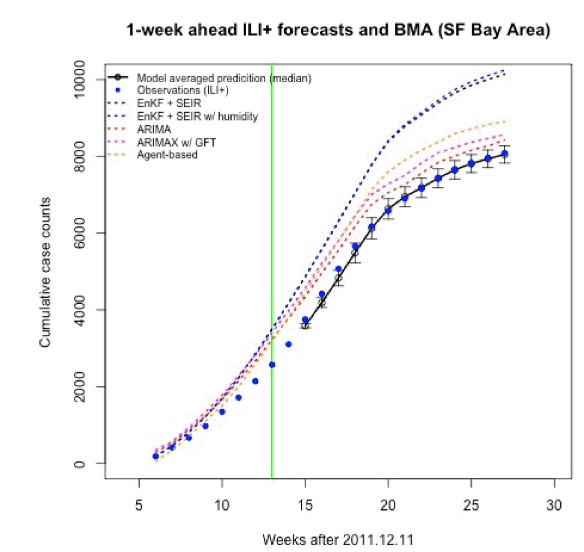
The Models

- Ensemble of 5 models
- 2 SEIR models driven by EnKF, assimilating ILI+; one also assimilates humidity
- 1 ARMA and 1 ARMAX model, also fitted to ILI+; ARMAX also uses Google Flu Trends
- 1 network model

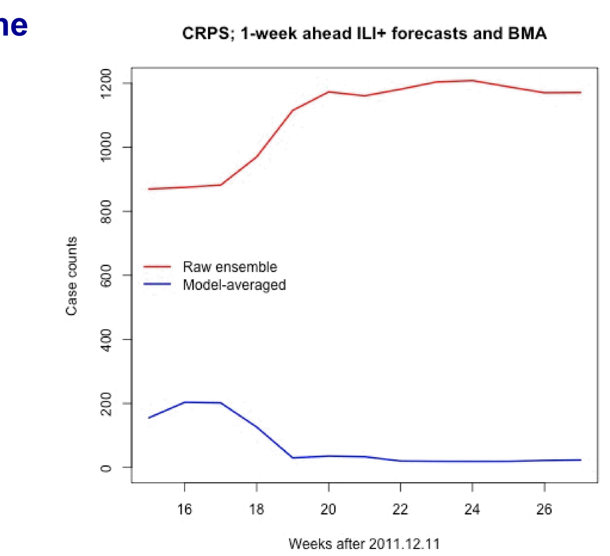


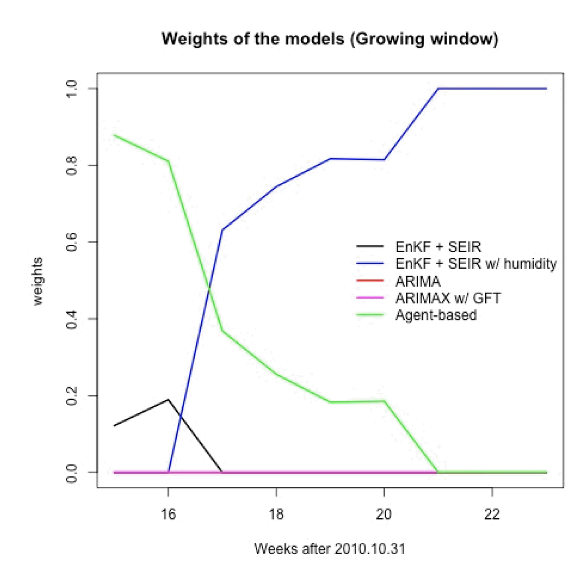
RESULTS

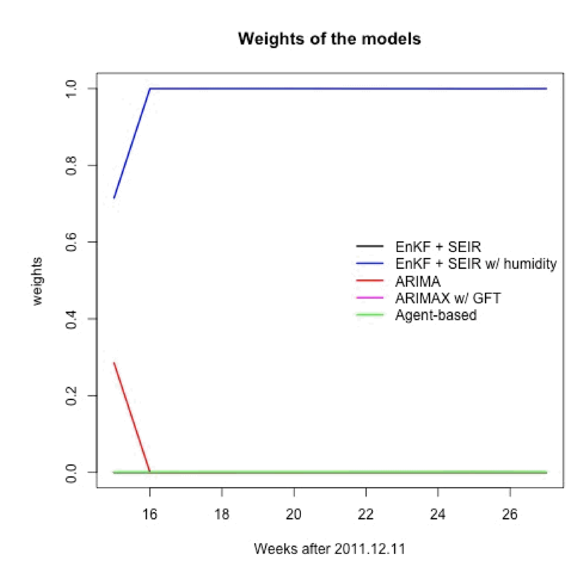




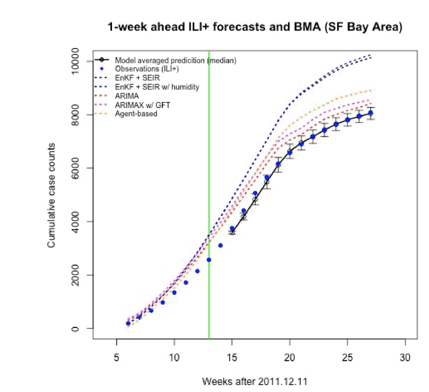
- BMA forecast much closer to the observed data than the individual models
- Prediction uncertainty (error bars) much tighter than the scatter in the ensemble
- Continuous Rank Probability Score (CRPS) of BMA-ed ensemble smaller than raw ensemble
- Different models get up-weighted during different flu

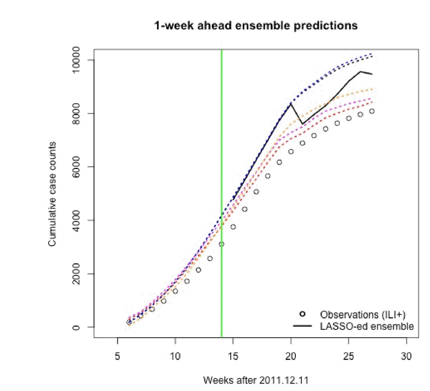


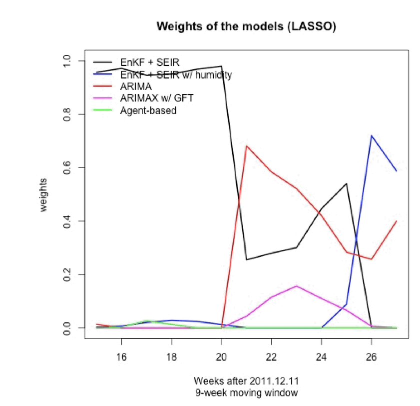




BMA OR SHRINKAGE?







- BMA-ed ensemble fits the data better than the one from shrinkage
- Shrinkage only removes 2 models
- Poor performance could be caused by lack of data to get a good λ

CONCLUSIONS

- We have a preliminary method based on BMA to reconcile conflicting model predictions
- BMA performs better than estimating the weighted ensemble via shrinkage
- No model dominates; certain models perform better than others in a given season
- Simpler models do not necessarily perform better, even during the early epoch of an outbreak when data is limited
- Relative importance of a model can change as a flu season progresses.

Acknowledgements

The project is supported by DTRA contract CBCall14-AMD3-CBA-06-2-0029.

References

- A. E. Raftery et al, "Using Bayesian Model Averaging to Calibrate Forecast Ensembles", Monthly Weather Review, 133.5, 2005: 1155-1174.

For additional information, please contact:

Jaideep Ray, Sandia National Laboratories, jairay@sandia.gov

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.