Statistical Modelling of Infectious Diseases: Influenza and the "Next Disease"

Shannon Gallagher, Lee Richardson, Sam Ventura, Ryan Tibshirani, Roni Rosenfeld†, & Bill Eddy
Department of Statistics, † Dept. of Machine Learning, Carnegie Mellon University

We want to improve disease forecasting

Better statistical forecasting includes:

- Understanding of underlying **assumptions**
- Incorporation of **high-quality data**
- Attention to **variance** of forecasts

Our work:

- Empirical Bayes model to forecast the flu
- Exploring agent-based models (ABMs)
- Visualizing the spread of disease

Disease is costly

For influenza in the US alone every year,

- Tens of thousands of lives are lost
- Nearly **$80B** in healthcare costs
- 31.4 million outpatient visits

Improved forecasting can alleviate these costs!

Forecasting the flu in the US

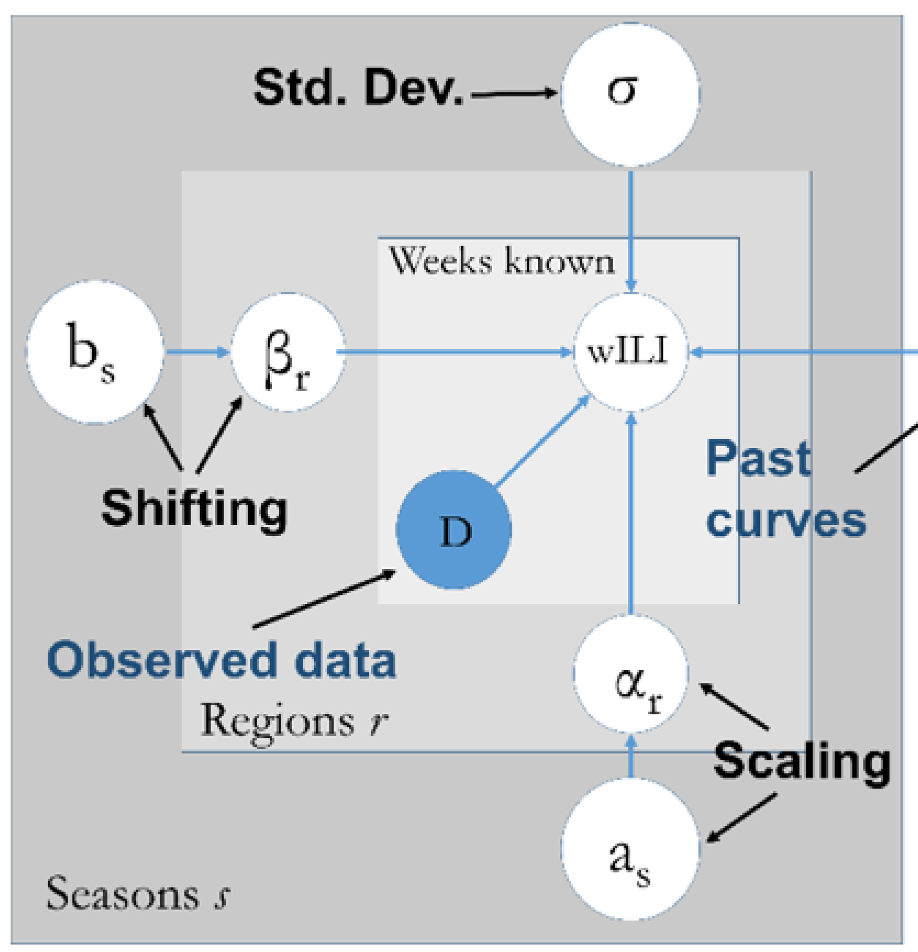
Goal: Forecast **wILI** (weighted influenza-like illness),

- For every week in the season (with new data each week)
- For 10 regions in the US
- Emphasis on peak week and peak wILI

Idea: A new flu season is going to look like a past one, with some scaling and shifting, *and* regions are dependent on one another

The model: Empirical Bayes with Regional Effects

Figure: Illustration of the model, consisting of seasonal and regional parameters. New data is incorporated each week.



The model: $wILI_t^{(r,s)} = [a_s \cdot \alpha_r] \cdot f(t - b_s - \beta_r) + \epsilon_t$, with $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$, for $r = 1, 2, \ldots, R$ and $s = 1, \ldots S$ and uniform priors for the parameters.

Forecasting the flu in the US (continued)

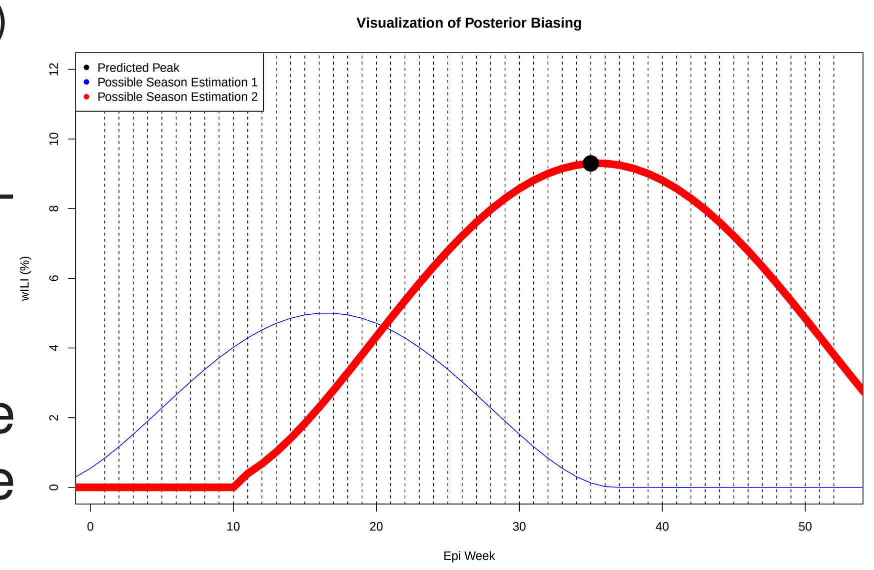
EB: Empirical Bayes (let $\alpha_r, \beta_r = 1$ for all r) (Brooks et al., 2015)

EBR: Empirical Bayes with regional effects

Problem: EBR is intractable and we are forced to make approximations for β_r

Solution: New approach: posterior biasing (**PB**)

Figure: Illustration of PB. The black dot is the peak. The blue curve is given less weight than the red one because the curve goes through the predicted peak.



Results: PB yields improved results, while EBR suffers from the approximation we made

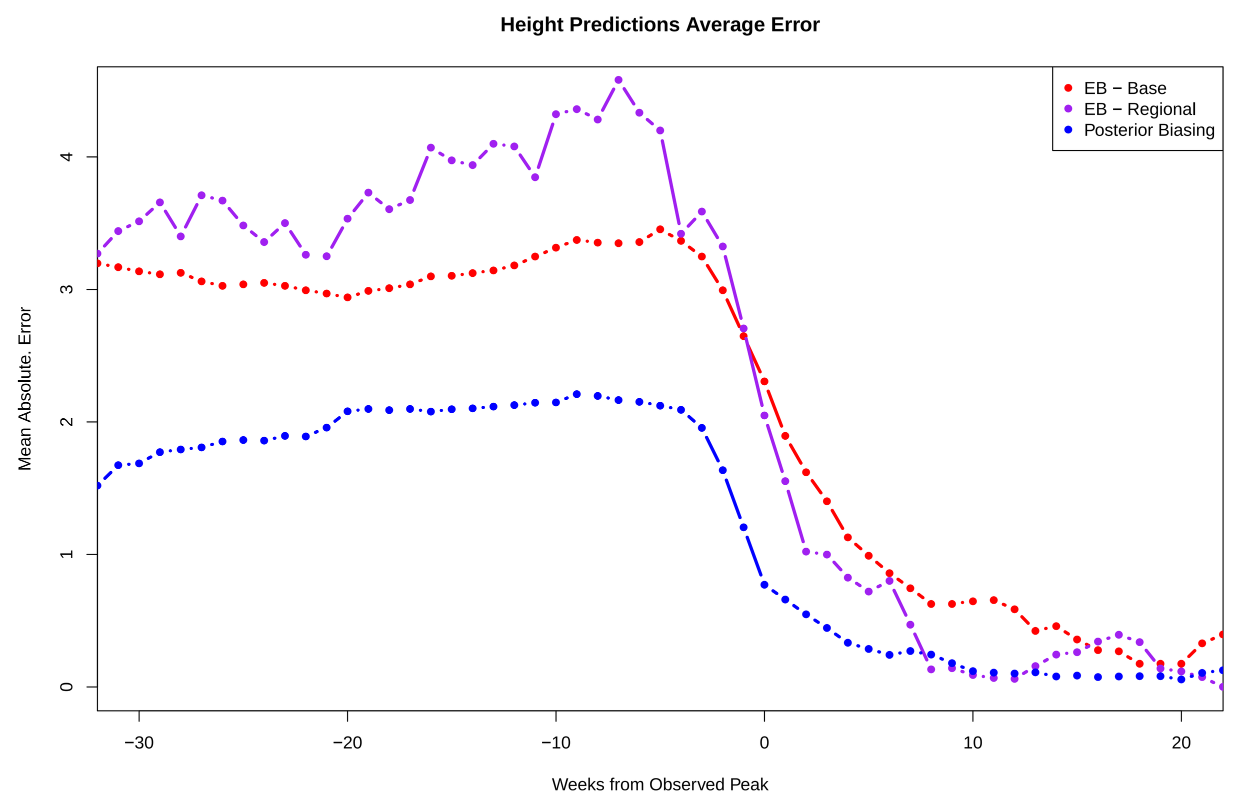


Figure: Leave one-season-out cross-validation (CV) of the three models. We calculate the CV for each week from the observed week as we expect better results the closer we are to the peak week

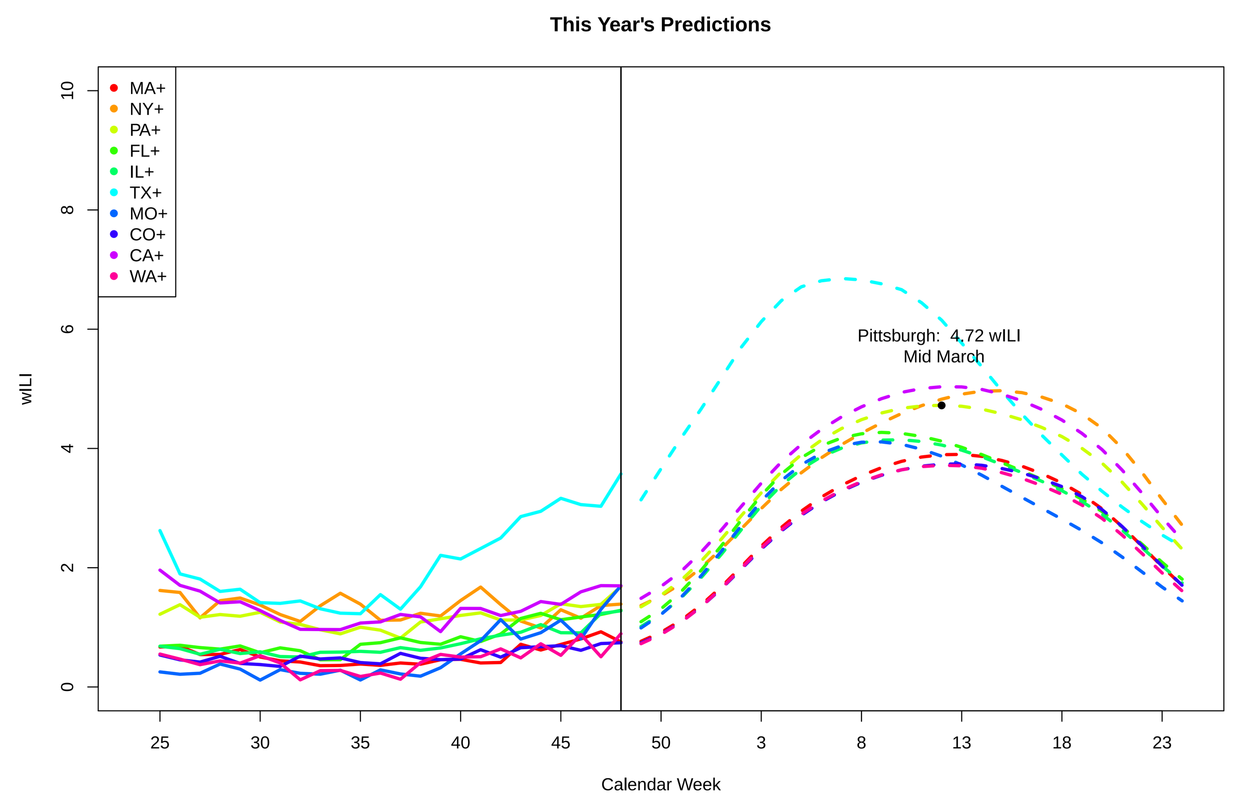
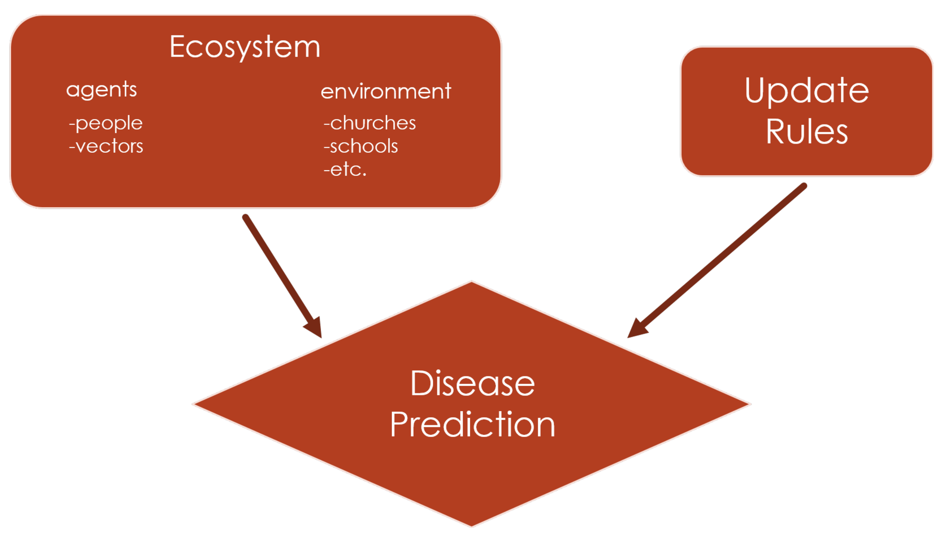


Figure: Current predictions for the (new) 2015-2016 Flu season with 25 weeks observed. The Texas region is predicted to be effected more intensely and sooner than the other regions

The "next" disease and agent-based models (ABMs)

ABMs are a viable way to forecast new diseases (e.g. Dengue, Ebola, Zika, the "next" one)



For new infectious diseases, we have

- Little data
- Less knowledge
- Frenzied awareness
- Few if any models

Solution: Simulate the spread of disease using ABMs!

SPEW: Synthetic Pop. & Ecosystems of the World

ABMs require **synthetic ecosystems** as input!

SPEW has generated high-quality input for ABMS:

- nearly 5 billion human agents
- 70+ countries and counting!
- automatic diagnostic reports

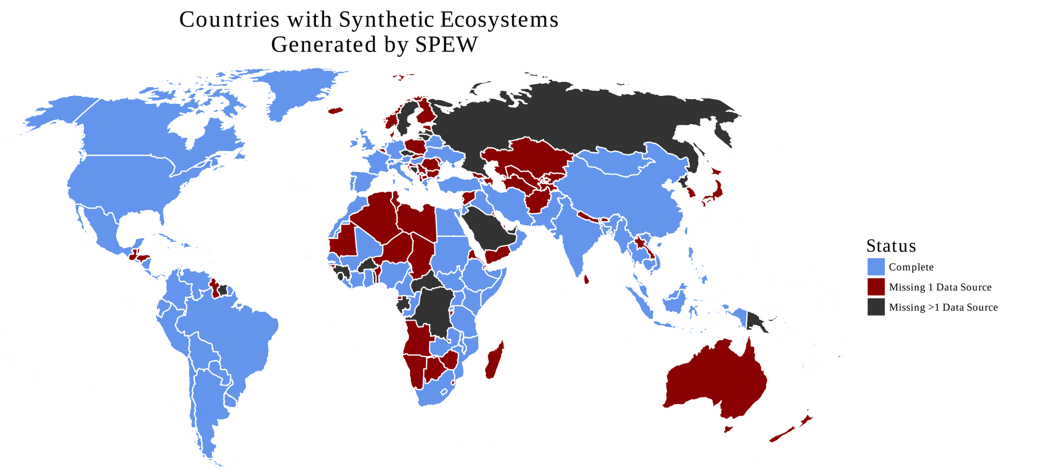


Figure: Synthetic ecosystems available at `epimodels.org`

SPEW View

`shiny.stat.cmu.edu:3838/sgallagh/spewview/`

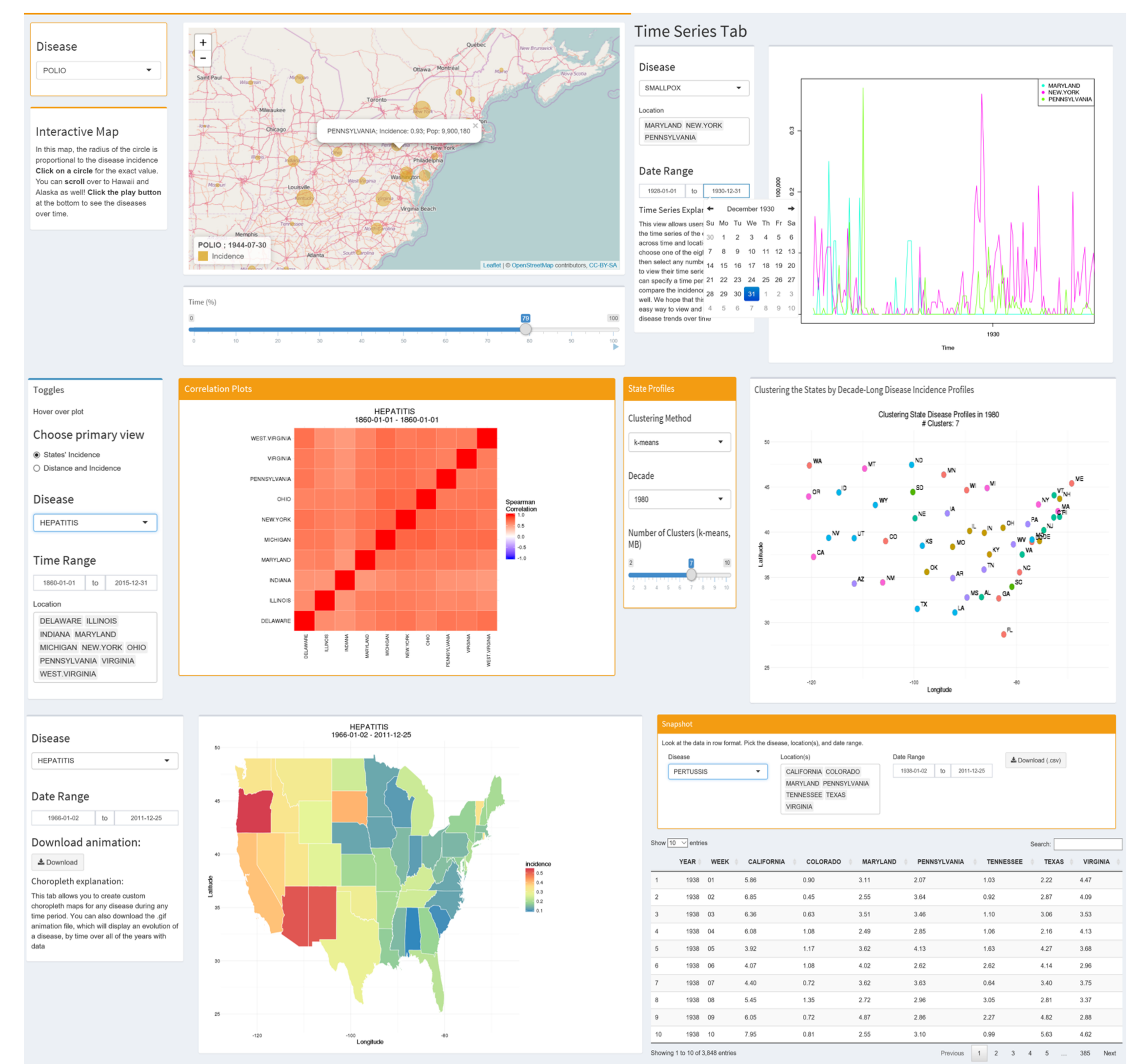


Figure: SPEW View. Data visualization of historical diseases in the US. **1st place** in Pittsburgh Supercomputing Center Public Health Hackathon

Acknowledgments

We would like to thank the Mihael Serban Memorial Endowment who made this poster possible. Additionally, we would like to thank the MIDAS Informatics Systems Group whom with NIH/NIGMS Grant 1 U24 GM110707-01. This work made extensive use of the Olympus Computing Cluster at the Pittsburgh Supercomputing Center (PSC).

MIDAS PSC