COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

Forecasting Infectious Disease Outbreaks

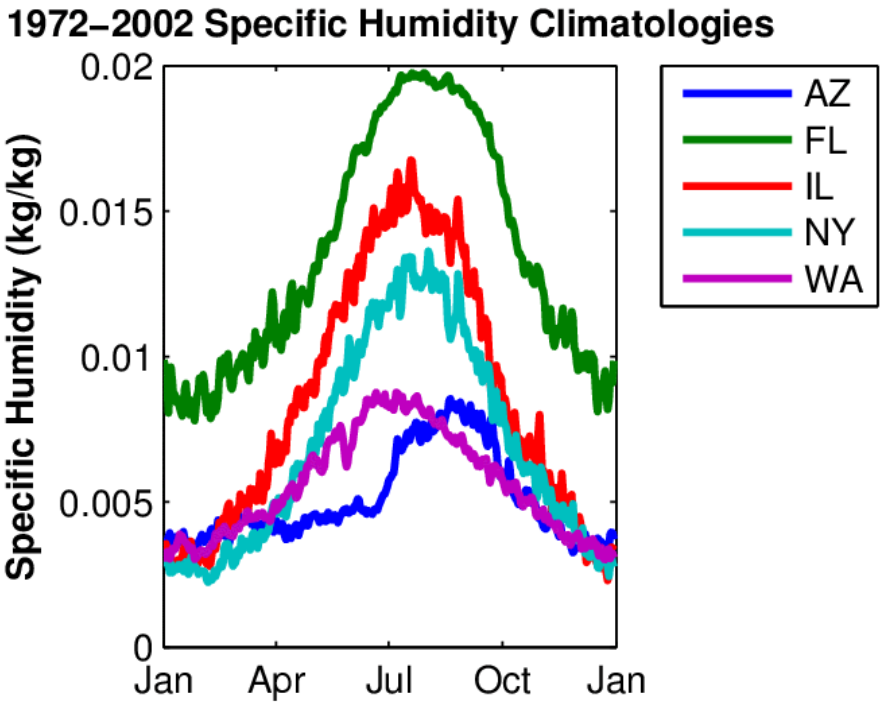
Jeffrey Shaman

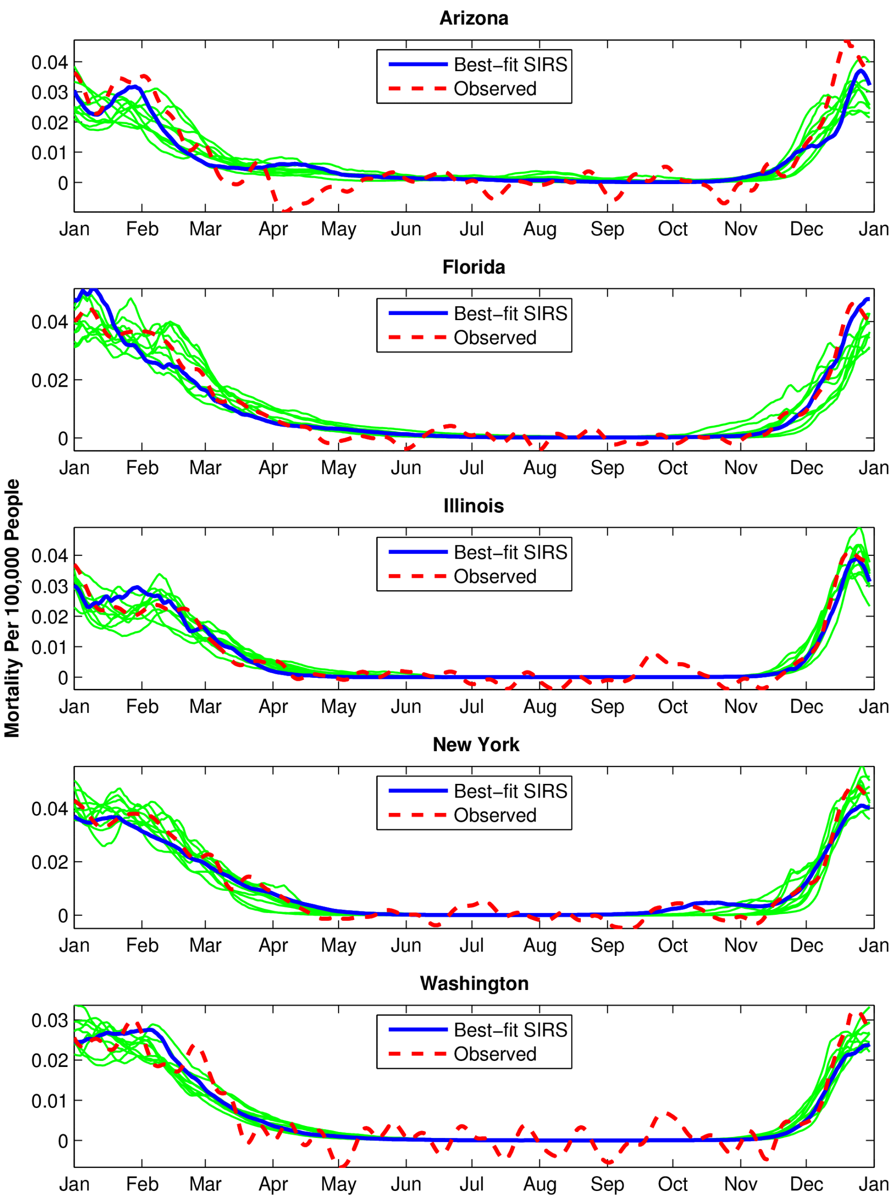
8 June 2016

Why Forecast Infectious Diseases?

- Infectious disease patterns continually shift
- Within outbreak response to infectious diseases is principally reactive (based on ongoing surveillance)
- Accurate, reliable forecasts with sufficient lead times would provide greater opportunity to plan adaptive mitigation and control efforts

Modeling the Seasonal Cycle of Influenza

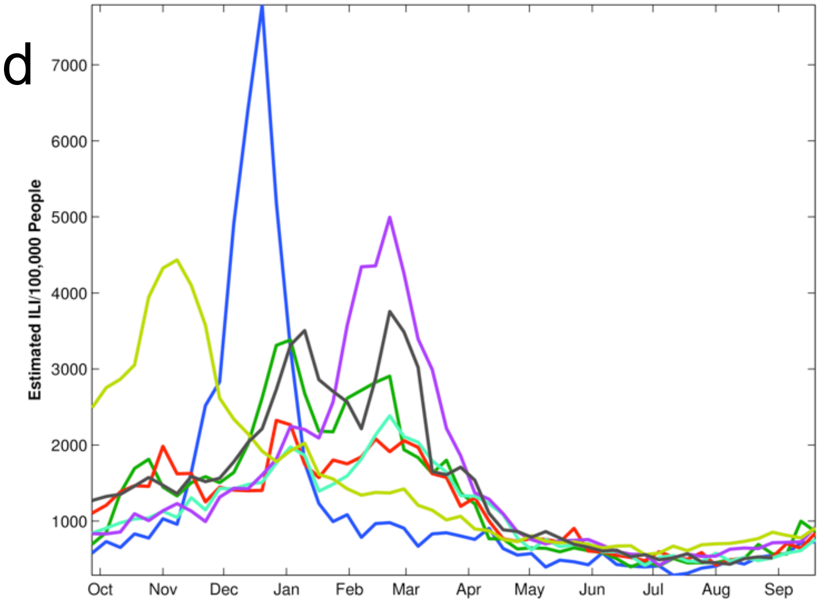




Can We Predict Individual Outbreaks?

- Seasonal flu dynamics are nonlinear and irregular

- Outbreaks, though in winter, vary enormously from year-to-year

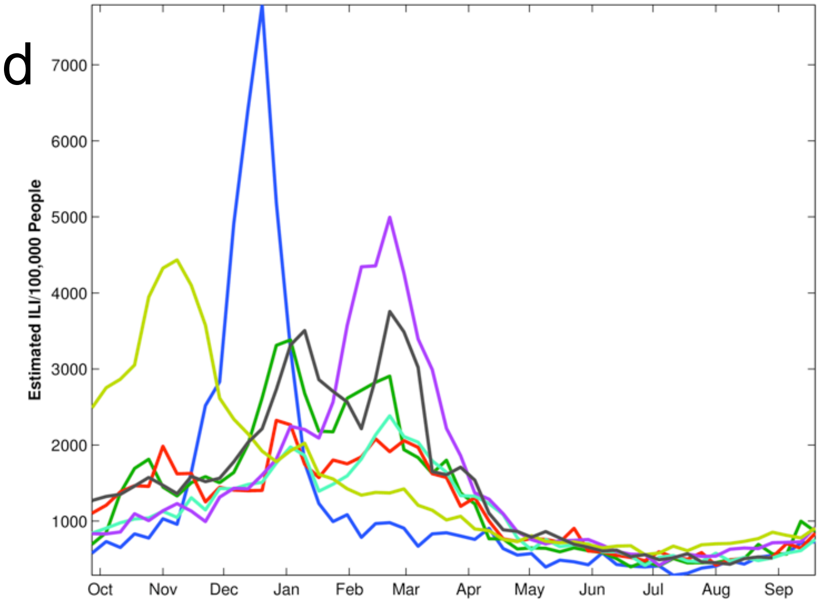


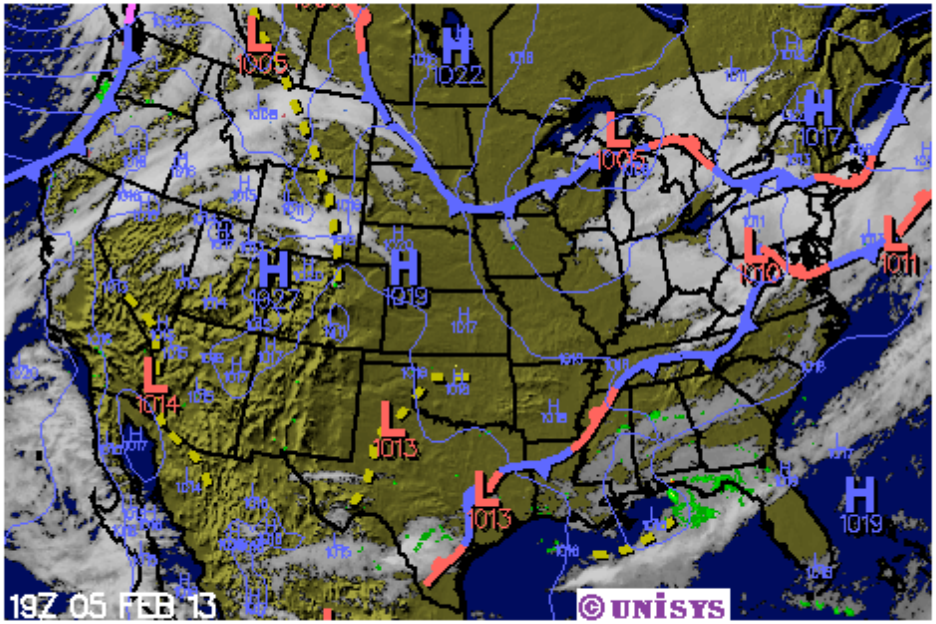
Can We Predict Individual Outbreaks?

- Seasonal flu dynamics are nonlinear and irregular

- Outbreaks, though in winter, vary enormously from year-to-year

- There are other systems with similar issues that are predicted

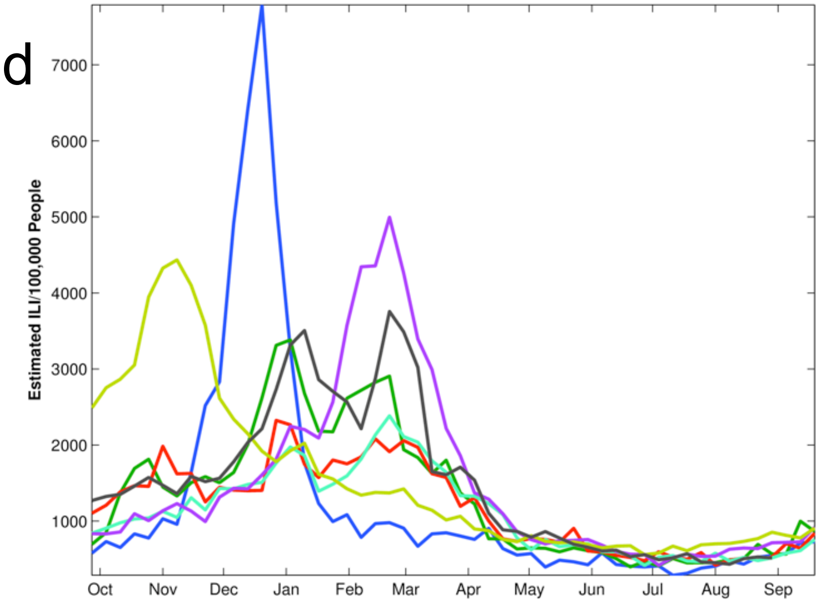




Our Forecasting Approach

• Seasonal flu dynamics are nonlinear and irregular

• Outbreaks, though in winter, vary enormously from year-to-year



To predict influenza, we mimic strategies used in numerical weather prediction

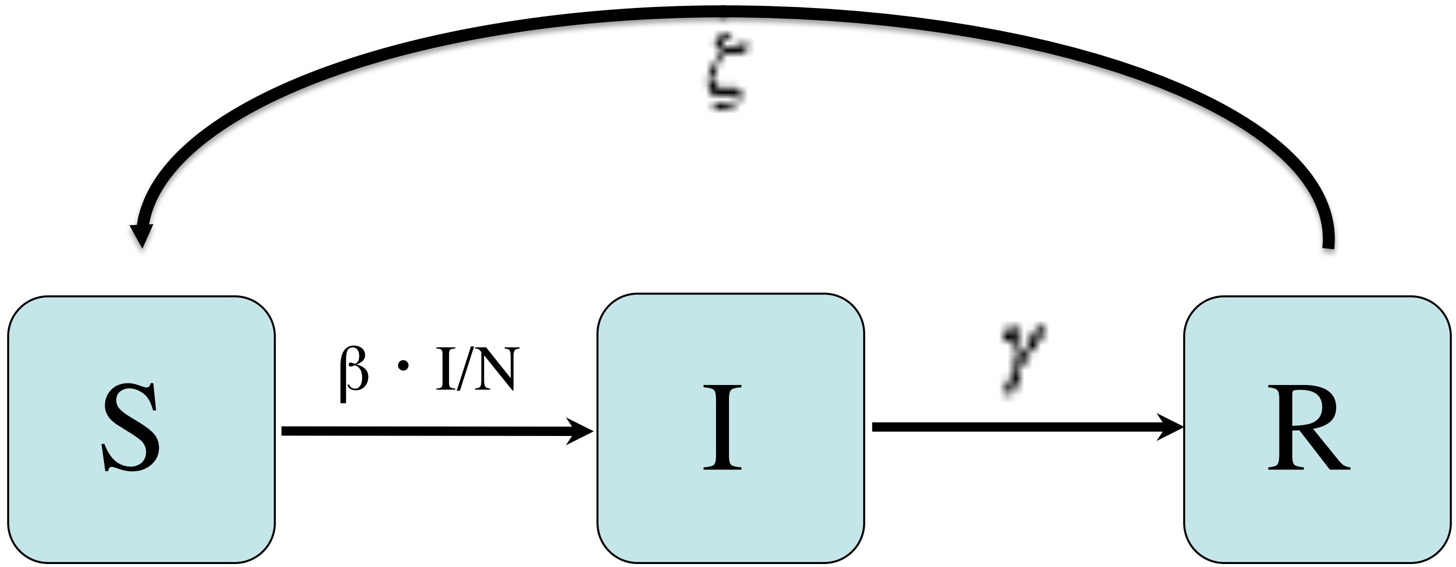
Requires 3 ingredients:

1) Observationally-validated model of influenza transmission dynamics

2) Real-time estimates of influenza infection rates (i.e. observations)

3) Data assimilation method to rigorously combine #1 and #2.

Humidity-forced SIRS Model



Here β is a function of observed daily specific humidity, a measure of absolute humidity

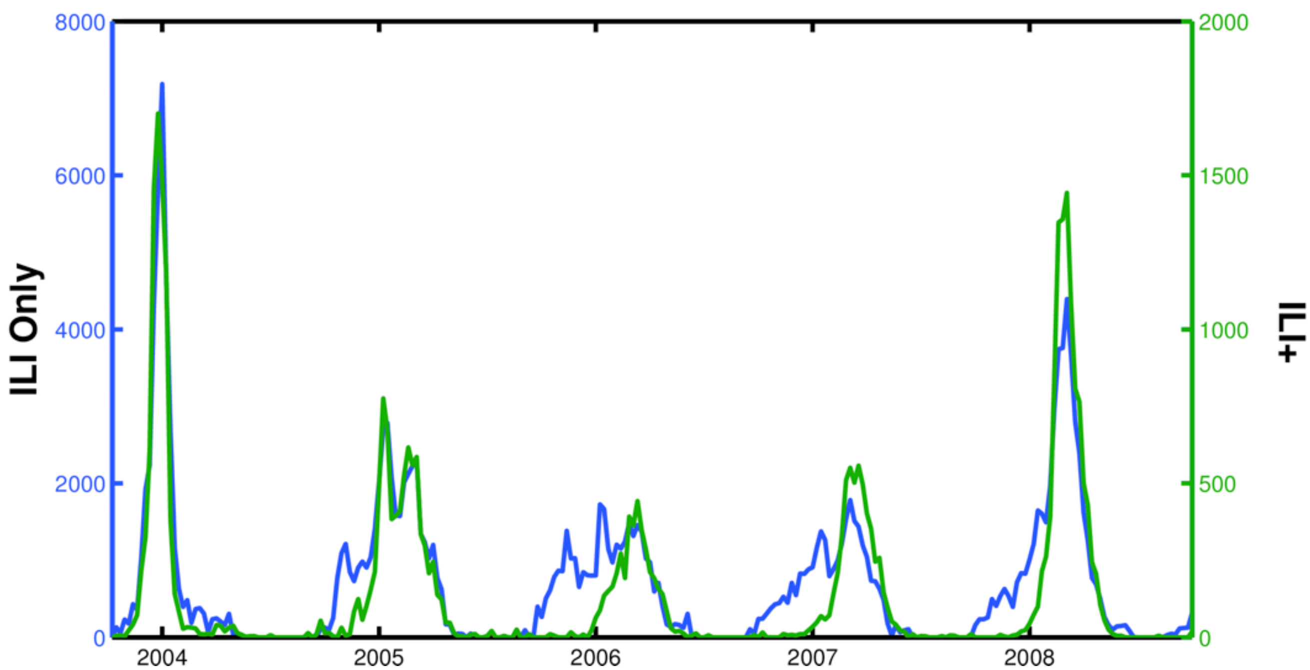
$$\frac{dS}{dt} = \frac{N - S - I}{L} - \frac{\beta(t)IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta(t)IS}{N} - \frac{I}{D}$$

Assessed fit to excess weekly P&I mortality via a conversion factor cases->lagged deaths

ILI+

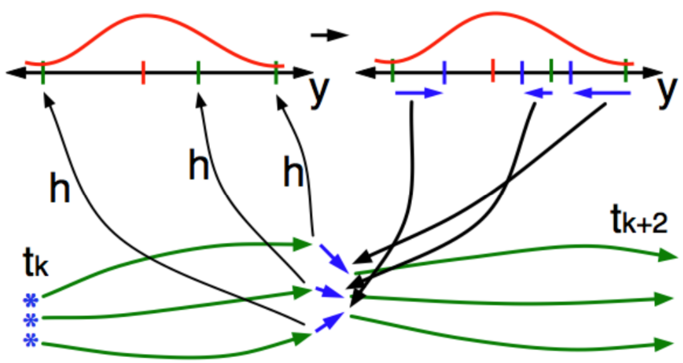
- For municipal forecasting, we often use a more specific estimate of influenza incidence
- We multiply municipal GFT ILI estimates by regional WHO/ NREVSS influenza positive test proportions
- The resulting metric (ILI+) eliminates signal from other respiratory infections, such as rhinovirus



Data Assimilation

Recursive (iterative) filtering of observations in a statistically rigorous fashion into an evolving model construct

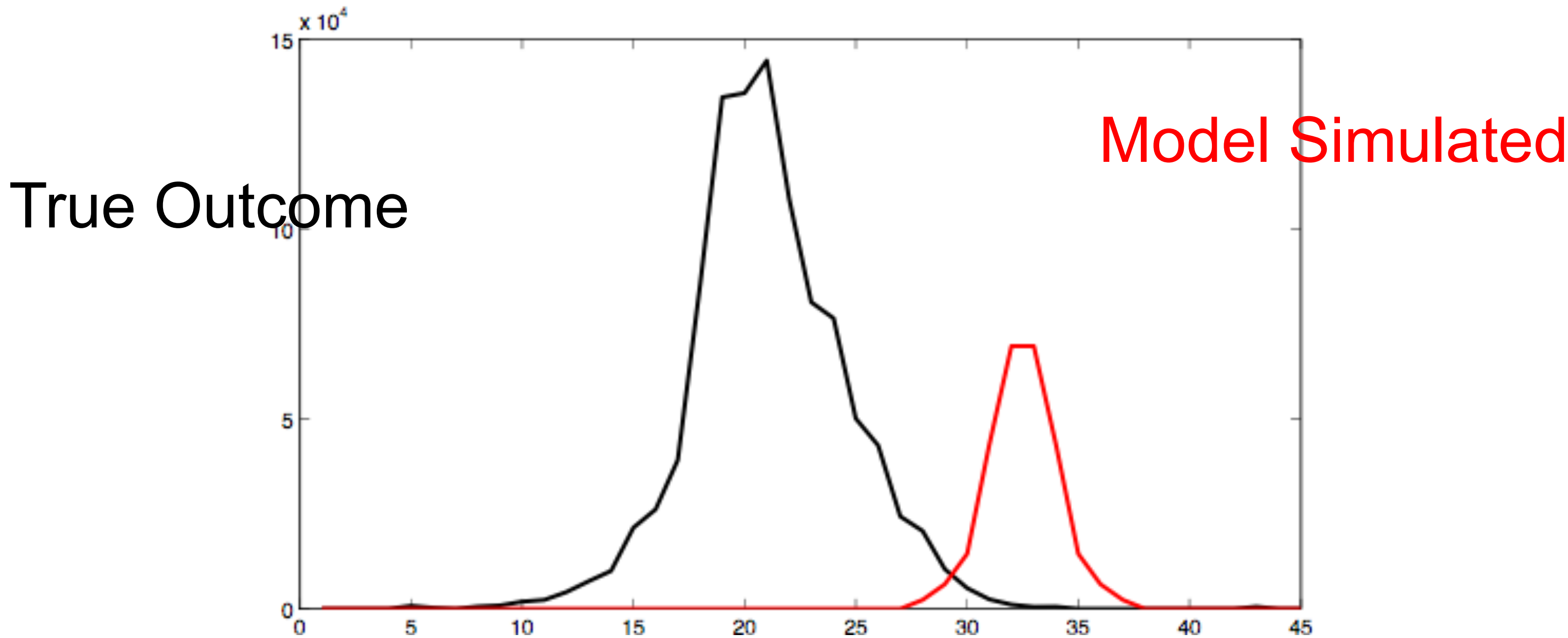
- Particle Filtering
- Kalman Filtering
- Variational Methods



Methods used in many disciplines, including numerical weather prediction where it is used to generate improved forecasts

Prior to Forecast: Training the Model

- Errors in the model structure, model parameters and initial model state amplify through time
- Left to its own devices the model forecast will deviate from reality



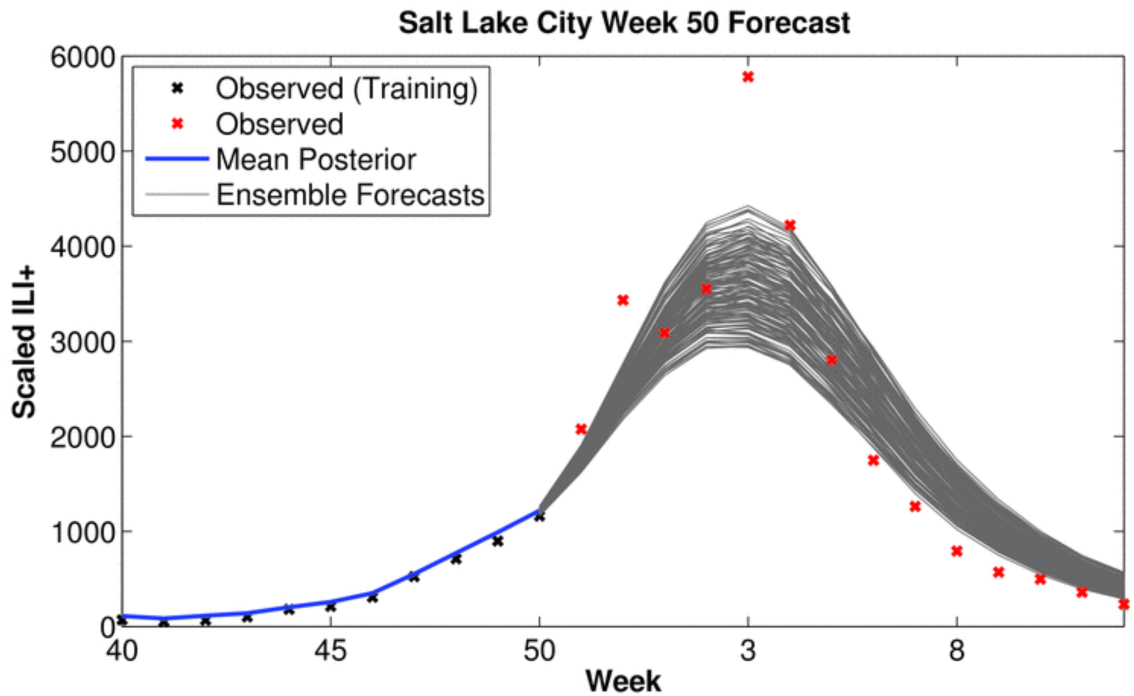
Prior to Forecast: Training the Model

- The real-time observations and data assimilation methods are used to recursively adjust and optimize the model
- By entraining observations up to the present the model forecast commences from a ‘truer’ starting point
- Because the unobserved variables and parameters have also been optimized, the model will evolve along more representative trajectories when integrated into the future
- The ensemble forecast itself is run following assimilation of the latest observation

Example Real-Time Forecast During 2012-2013

Forecasts (grey lines) made with an SIRS model

Model recursively trained using real-time observations (black ‘x’) and data assimilation methods up to the point of forecast (Week 50)



Observed estimates of influenza incidence that were in the future at the time of forecast are shown as red ‘x’.

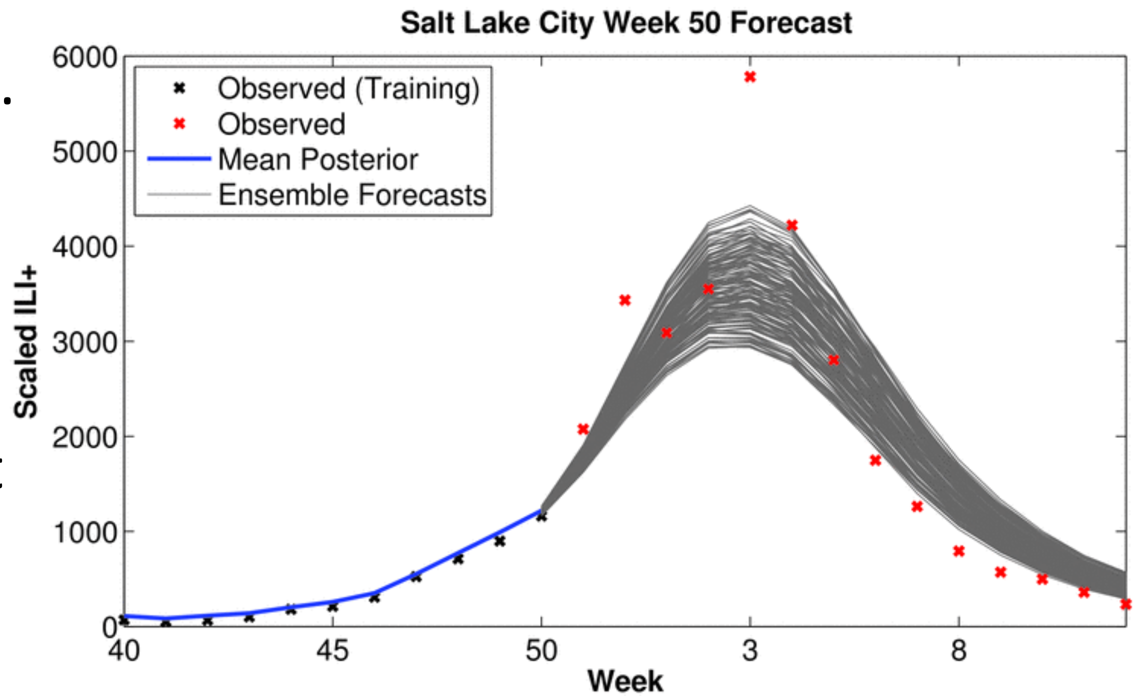
A Calibrated Forecast

Do not simply want to predict an outcome (e.g. the peak will occur in 5 week)

Want to know the certainty of the forecast as it is made

Is there a 90% chance the peak will occur in 5 weeks?

Is there a 20% chance?

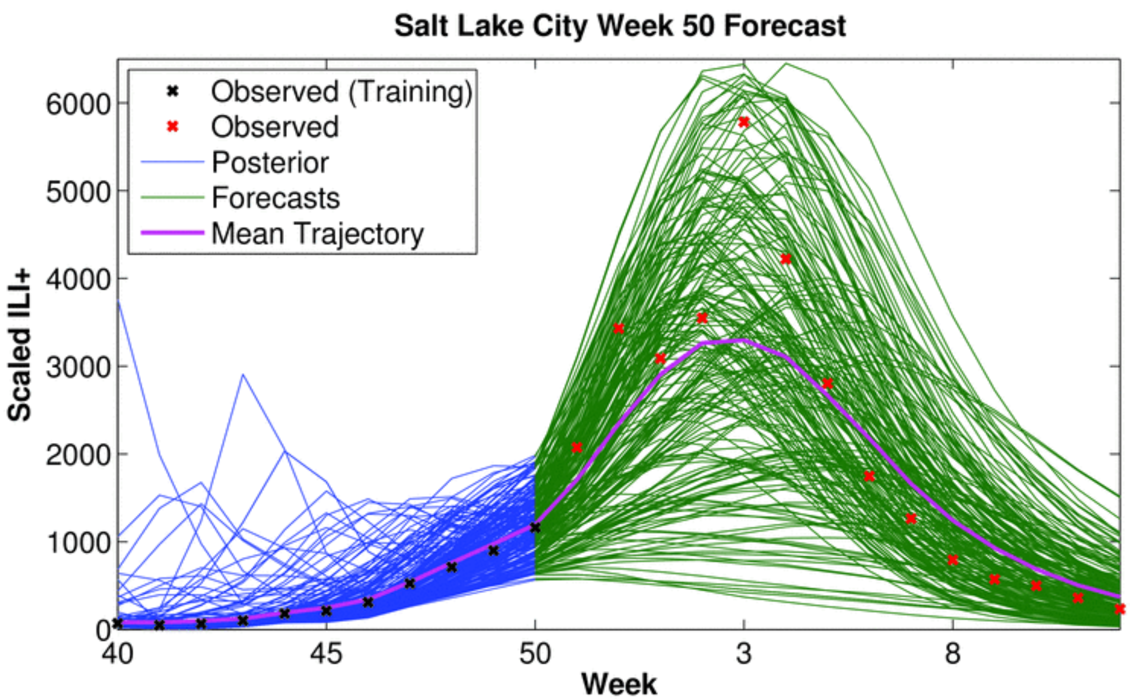


Accurate ascription of forecast certainty provides the public health user a much richer, more actionable prediction

A Calibrated Forecast

It turns out, we can use the spread of each ensemble of predictions to estimate the certainty of a forecast

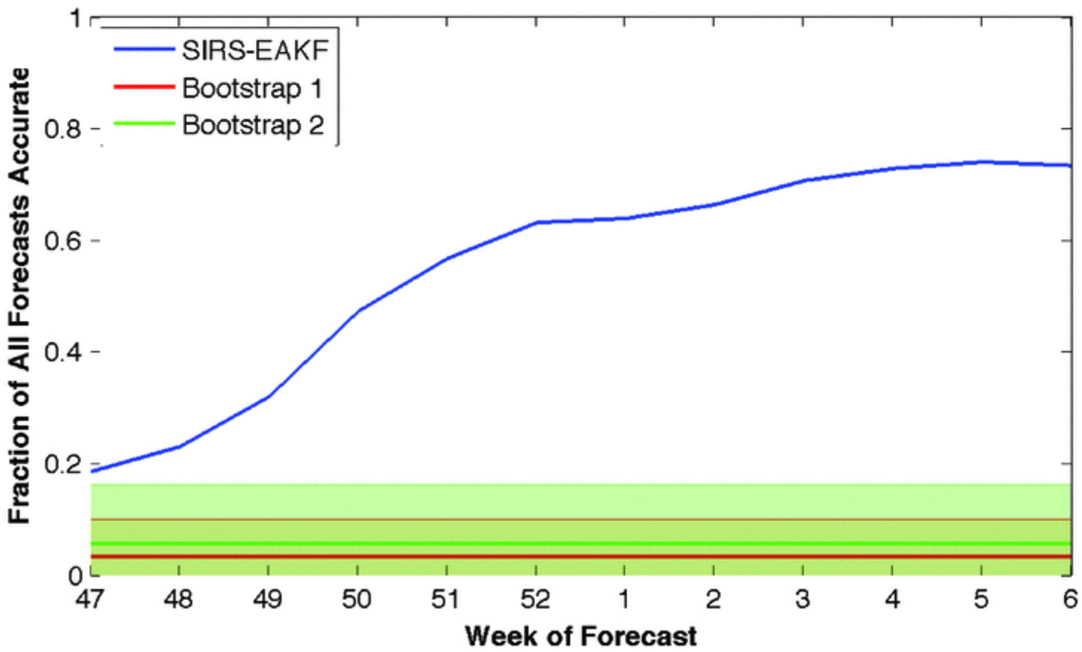
The relationship between that spread (variance) and accuracy for past forecasts can be used to calibrate forecasts made in real time



Above plot now shows the individual trajectories within a single ensemble forecast

Predicting Peak Timing

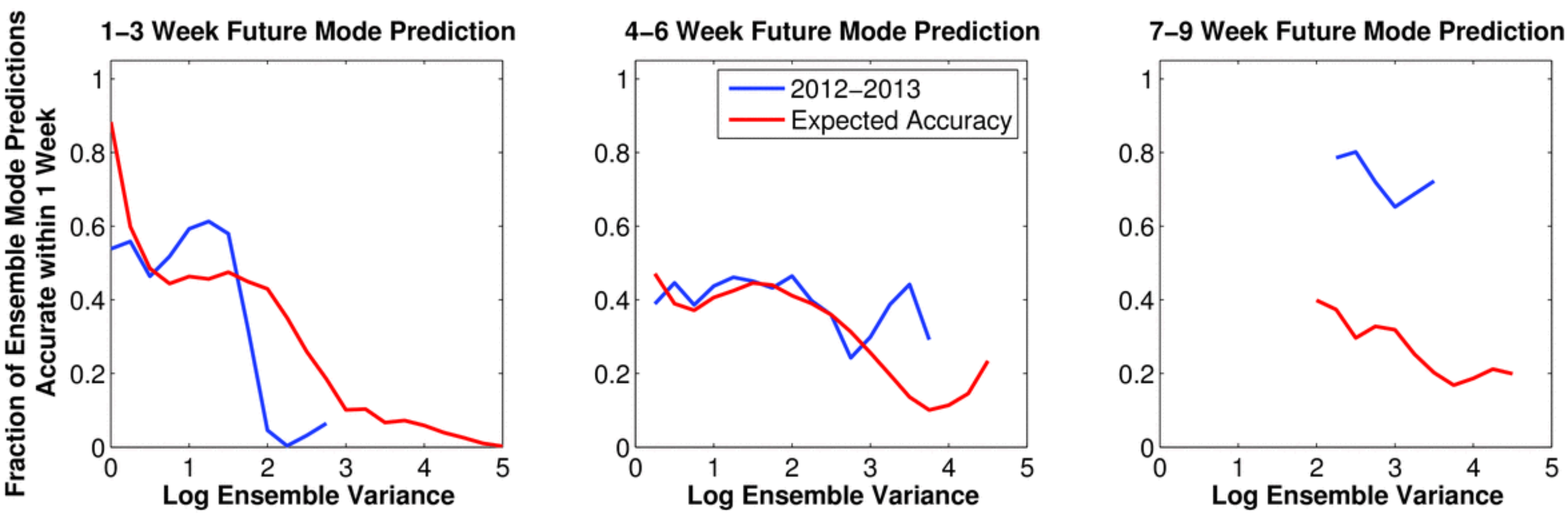
- A number of issues to be verified:
  - The accuracy of the forecasts -- by Week 52 of the 2012-2013 season 63% of forecasts for 108 cities were accurately forecast (84% of cities peaked Week 2 or later)



Predicting Peak Timing

- A number of issues to be verified:
  - The expected accuracy of the forecasts

Would hope that the forecasts gauged to be correct 80% of the time are correct 80% of the time (and the forecasts gauged to be correct 20% of the time are correct 20% of the time)



  - The forecast lead: Up to 9 weeks

Much Work Remains

- Can we build a more reliable forecast model?

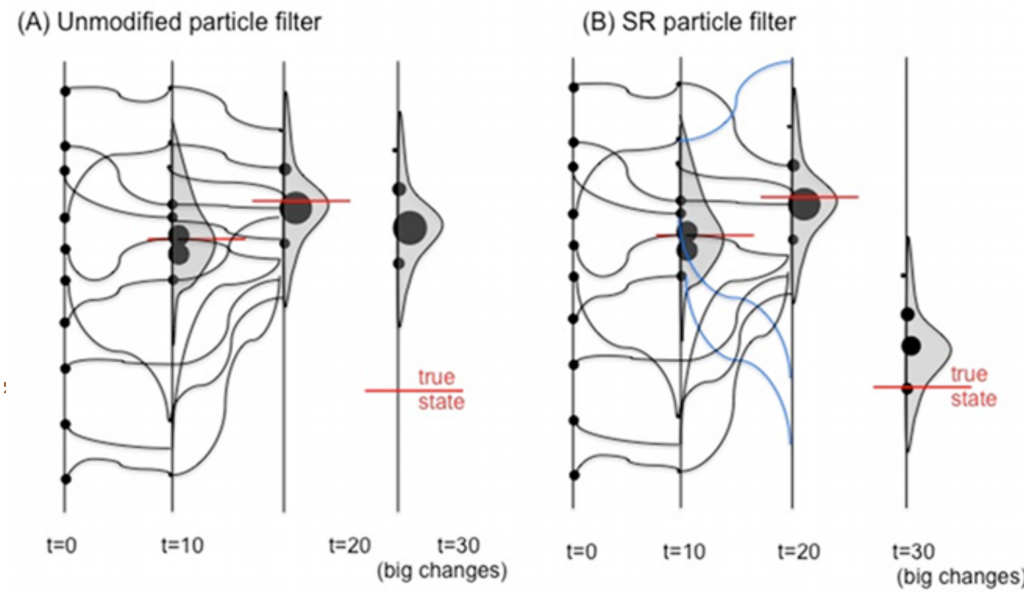
Testing Alternate Model Forms (age-stratified, stochastic v. deterministic, multiple strains, spatially explicit)

- Can we improve model optimization?

Testing and creating different data assimilation methods (ensemble filters, particle filters)

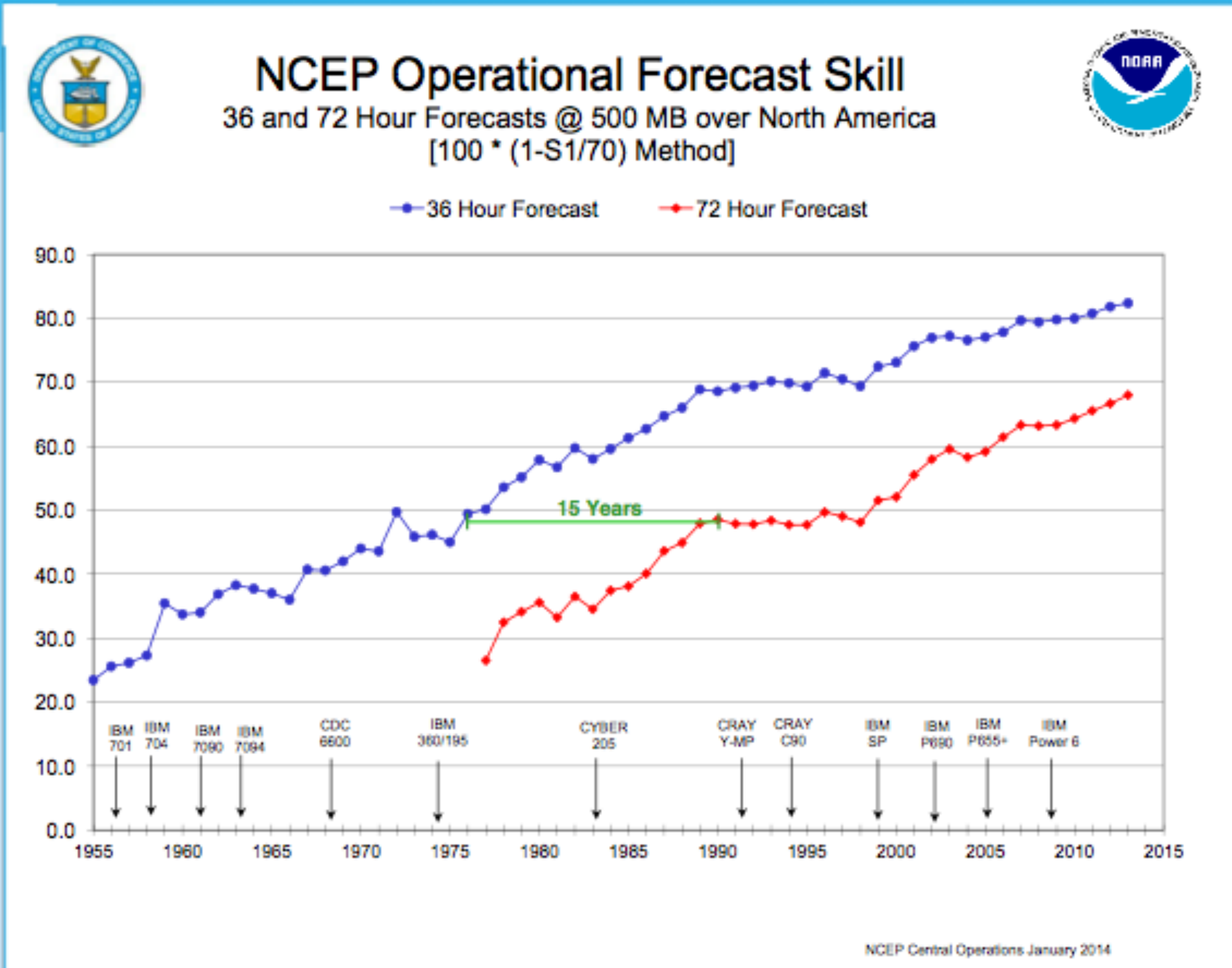
- Can we provide forecasts for local public health use?

Testing different observations of influenza (Google, CDC, Twitter, Wikipedia, WHO)



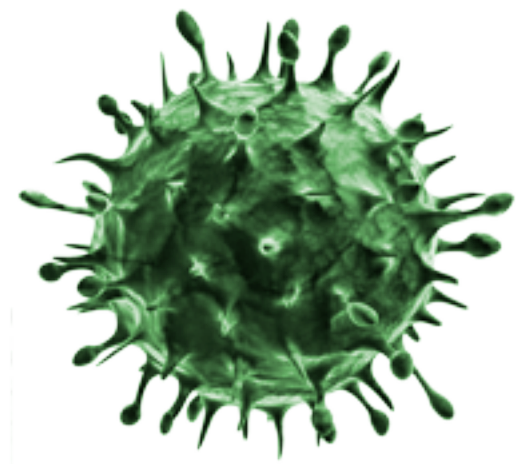
Yang and Shaman, 2014

Increasing skill



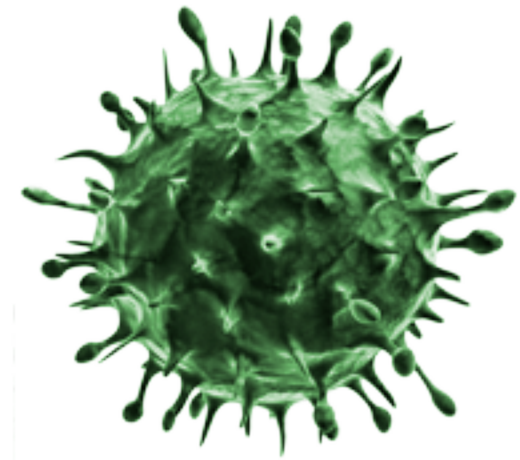
Possible Uses:
General Public

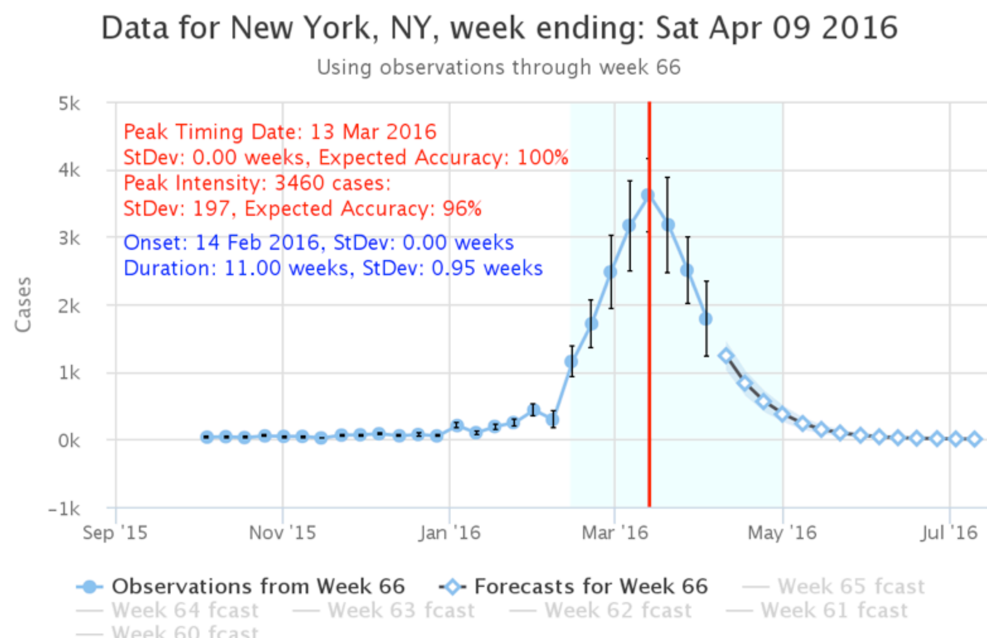
- Nightly Flu Forecast
- Lead Times for Peak Sufficient for Vaccination and Adaptive Immune Response/Antibody Production
- Awareness of Germs Circulating Locally may Affect Behavior -- stay home from work or keep kids home from school when sick, cancel play dates, etc.



Possible Uses:
Public Health Officials

- Distribution of vaccines, medicines and supplies to regions with more urgent need
- Inform school closure decisions in the event of a virulent outbreak
- Hospital resource and staffing management planning
- Timely Public Service Announcements





Collaborators

Alicia Karspeck, National Center for Atmospheric Research

Marc Lipsitch, Harvard School of Public Health

Wan Yang, Columbia University

James Tamerius, University of Iowa

Sasi Kandula, Columbia University

Haruka Morita, Columbia University

Alexandra Heaney, Columbia University

Sarah Kramer, Columbia University

Funders

NIH (NIGMS)/NSF (DMS) joint initiative to support research at the interface of the biological and mathematical sciences

MIDAS
Models of Infectious Disease Agent Study
Funded by the National Institutes of Health

bioCSL

DARPA

DEFENSE THREAT REDUCTION AGENCY
UNITED STATES OF AMERICA

BARDA
BIOMEDICAL ADVANCED RESEARCH AND DEVELOPMENT AUTHORITY