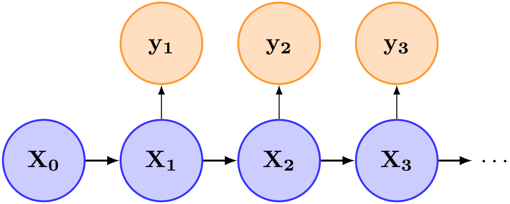
Hierarchical modeling and data fusion

Jarad Niemi and Nicholas Michaud

Iowa State University

August 31, 2016

State-space model



Let

- X_t represent the latent state of the outbreak at time t, e.g. the proportion of people with ILI
- y_t represent the observed data at time t

The arrows describe the

$$\begin{array}{ll} \text{evolution equation:} & x_t \sim p(x_t|x_{t-1}, \theta) \\ \text{observation equation:} & y_t \sim p(y_t|x_t, \phi). \end{array}$$

Evolution equation

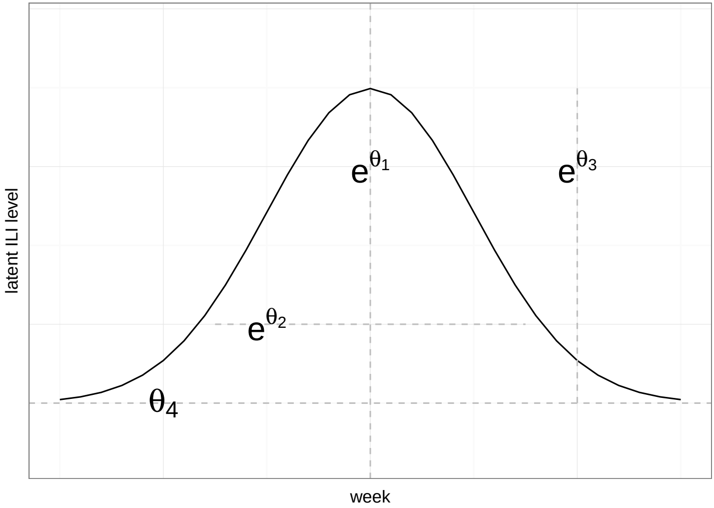
- Deterministic models, e.g.

$$x_t = f(t;\theta) = \theta_4 + e^{\theta_3}\phi\left(\frac{t - e^{\theta_1}}{e^{\theta_2}}\right)$$

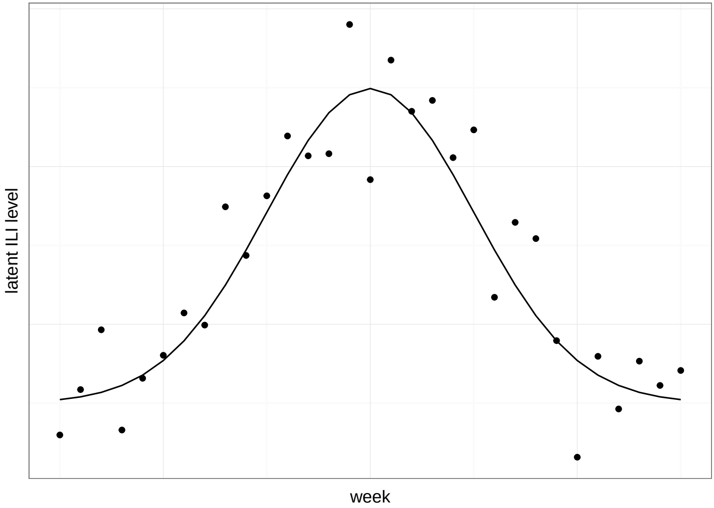
- Stochastic models

$$\begin{aligned}
x_t &= (s_t, i_t, r_t) \\
s_t &= s_{t-1} - \overrightarrow{si}_t \\
i_t &= i_{t-1} + \overrightarrow{si}_t - \overrightarrow{ir}_t \\
r_t &= r_{t-1} \qquad\quad + \overrightarrow{ir}_t \\
\overrightarrow{si}_t &\sim Po(\theta_1 s_t i_t) \\
\overrightarrow{ir}_t &\sim Po(\theta_2 i_t)
\end{aligned}$$

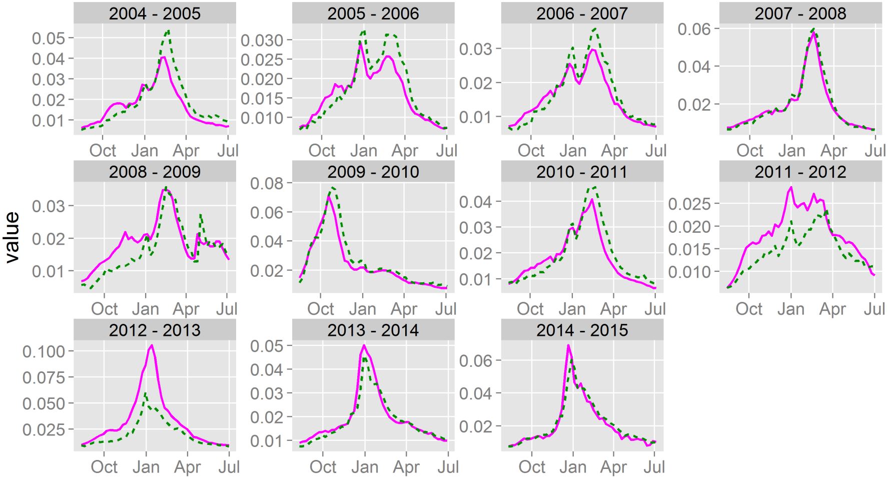
Example evolution



Example evolution



ILI-Net and Google Flu Trends



Observation models

CDC ILI Net data (I_t is logit of proportion):

$$I_t = x_t + e_t, \quad e_t \stackrel{ind}{\sim} N(0, \sigma_I^2)$$

Data fusion:

- Google Flu Trends
- Twitter symptom tweeting
- Wikipedia editing

Biased data (generically B_t) model:

$$B_t = \beta_t + x_t + w_t, \quad w_t \stackrel{ind}{\sim} N\left(0, \sigma_B^2\right)$$

Possible models for bias:

$$\begin{aligned} \beta_t &= \rho\beta_{t-1} + v_t, & v_t &\stackrel{ind}{\sim} N(0, \tau^2) \\ \beta_t &= \rho[I_{t-1} - x_{t-1}] + v_t, & v_t &\stackrel{ind}{\sim} N(0, \tau^2) \end{aligned}$$

Hierarchical model

Each season has its own parameters, but we want to borrow information across seasons.

Season-specific model:

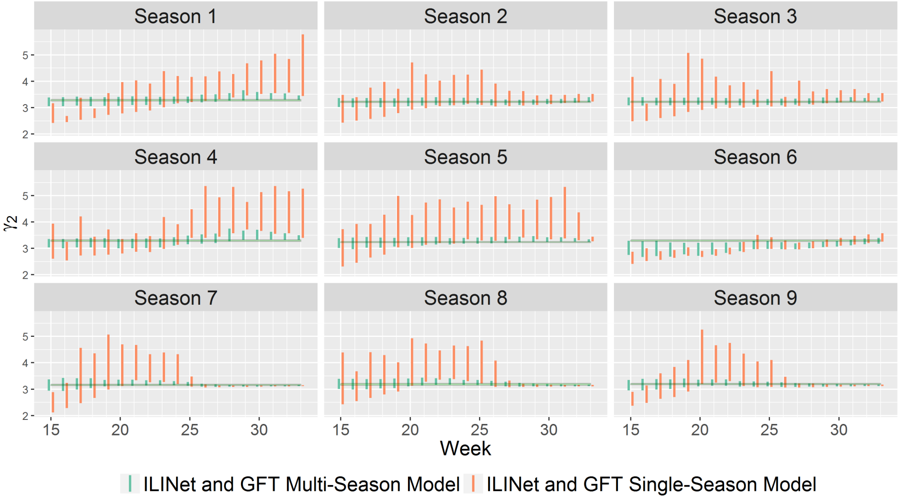
$$
\begin{aligned}
y_{s,t} = \left[\begin{array}{c} I_{s,t} \\ B_{s,t} \end{array}\right] &\stackrel{ind}{\sim} N\left(\left[\begin{array}{r} x_{s,t} \\ \beta_{s,t} + x_{s,t} \end{array}\right], \left[\begin{array}{cc} \sigma^2_{s,I} & 0 \\ 0 & \sigma^2_{s,B} \end{array}\right]\right) \\
\beta_{s,t} &\stackrel{ind}{\sim} N(\rho_s \beta_{s,t-1}, \tau^2_s) \\
x_{s,t} &= \theta_{s,4} + e^{\theta_{s,3}} \phi\left(\tfrac{t - e^{\theta_{s,1}}}{e^{\theta_{s,2}}}\right)
\end{aligned}
$$

Hierarchical model:

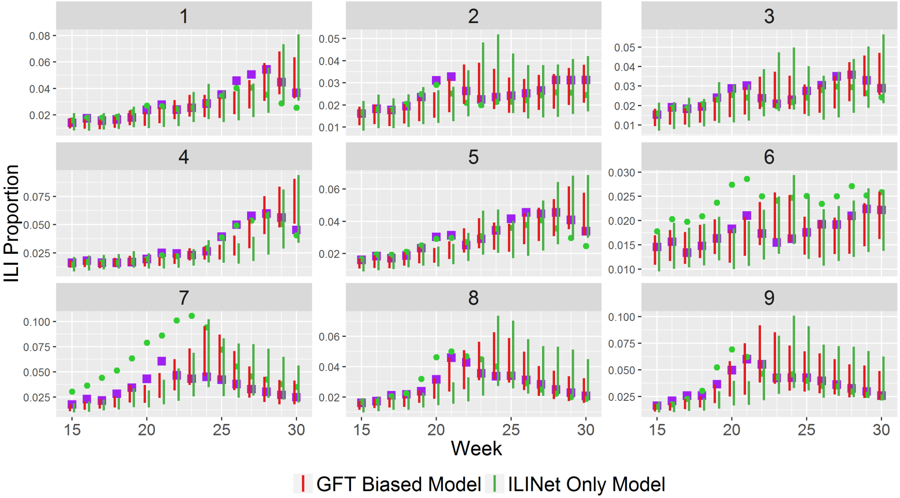
$$
\begin{aligned}
\sigma^2_{s,\cdot} &\stackrel{ind}{\sim} LN(\alpha_\cdot, \beta_\cdot) \\
\theta_s &\stackrel{ind}{\sim} N(\mu, \Sigma)
\end{aligned}
$$

The hierarchical model provides a data-based way to borrow information across seasons.

Estimating (log of) peak week



One- (three-) week forecasts



R packages and apps

```r
# This presentation
browseURL("https://github.com/jarad/CDC2016")                  # Code
browseURL("http://www.jarad.me/research/presentations.html") # Presentation

# NIMBLE
browseURL("http://r-nimble.org/")

# Convert back and forth from MMWR weeks
install.packages("MMWRweek")
library(MMWRweek)
MMWRweek(as.Date(Sys.time()))

# Visualize CDC data
shiny::runGitHub('NLMichaud/WeeklyCDCPlot')
browseURL("https://gallery.shinyapps.io/CDCPlot/") # data not updated
```