北海道大学
HOKKAIDO UNIVERSITY

感染症流行のリアルタイム予測
Real time forecasting of ID epidemics

Kobe, Riken

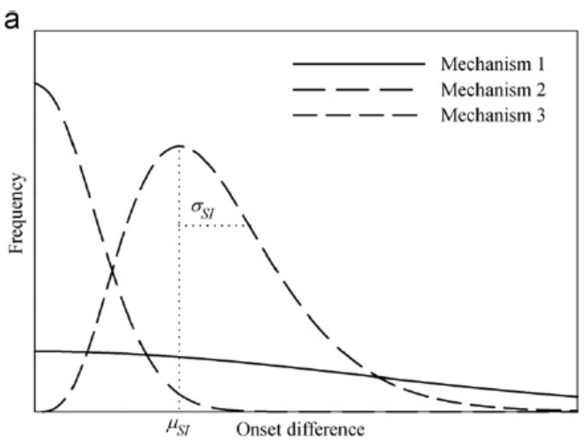
25 Sep 2017

Hiroshi Nishiura

Hokkaido University

nishiurah@med.hokudai.ac.jp

First meet with sequential Monte Carlo



$$f(d) = \pi_1(q_1(d) + q_1(-d)) + \pi_2(q_2(d) + q_2(-d)) + (1 - \pi_1 - \pi_2)(q_3(d) + q_3(-d))$$

$$q_2(d) = \int_0^\infty g(\tau)g(d+\tau)d\tau$$

$$q_3(d) = \int_0^\infty \int_0^\infty h(\tau,u)g(d-u+\tau)dud\tau$$

2008/9

on measles

Objective: decompose mechanisms and purify household transmission

n=4516

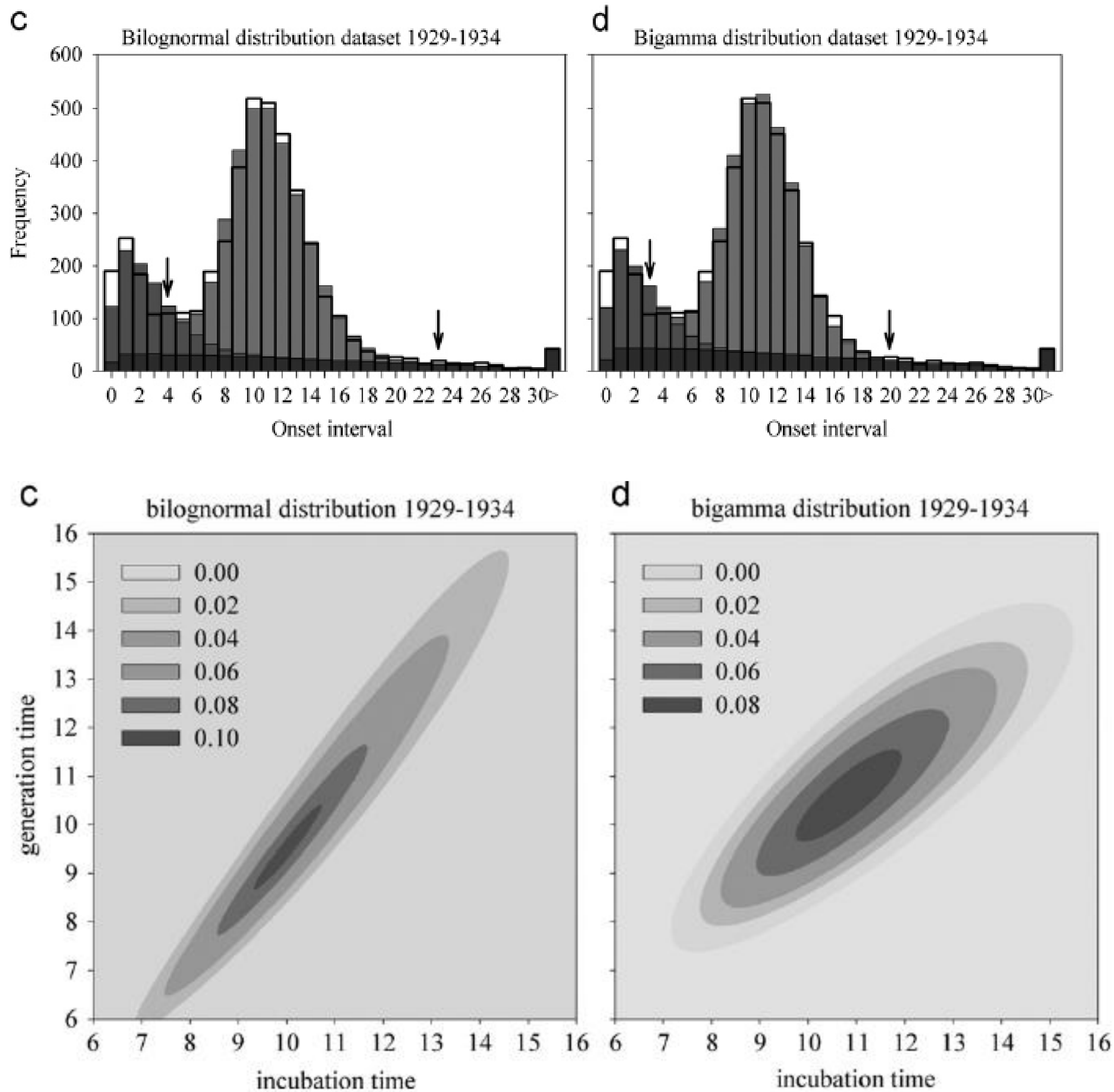
Implemented: Mathematica

MLE and classical Bayesian: challenging

ABC-PRC (Sisson et al. (2007))

- 2000 parameter sets from prior
- Second parameter sets: simulation + chi2 test with threshold ipsilon1
- Third: chi2 test with ipsilon2…

Klinkenberg & Nishiura, J Theor Biol 2011



Klinkenberg & Nishiura, J Theor Biol 2011

Forecasting-associated characteristics of IDs

1. Stochastic dependence structure
   Demographic stochasticity + sampling error
2. Several “acceptable” models and assumptions
   e.g. “SIR model” + “age-dependent heterogeneity”
3. Observed datasets are poor (感染症発生動向調査）
4. Practical demand: too much

Two kinds of ID prediction

1. Short term

You're in the middle of an epidemic (real time).

Say what will happen next in that epidemic

2. Long term

You're between epidemics.

Plan how effective school closures would be for an epidemic like H1N1-2009

Advice: long-term predictions

- Right level of structure in the model
- Prediction: likely quantitatively wrong

  reasons: system will change

  no info on future epidemic

  no idea how effective the intervention will be

"Scenario analysis" = "Not really a statistical problem"

Short term
SDS (Stochastic dep)

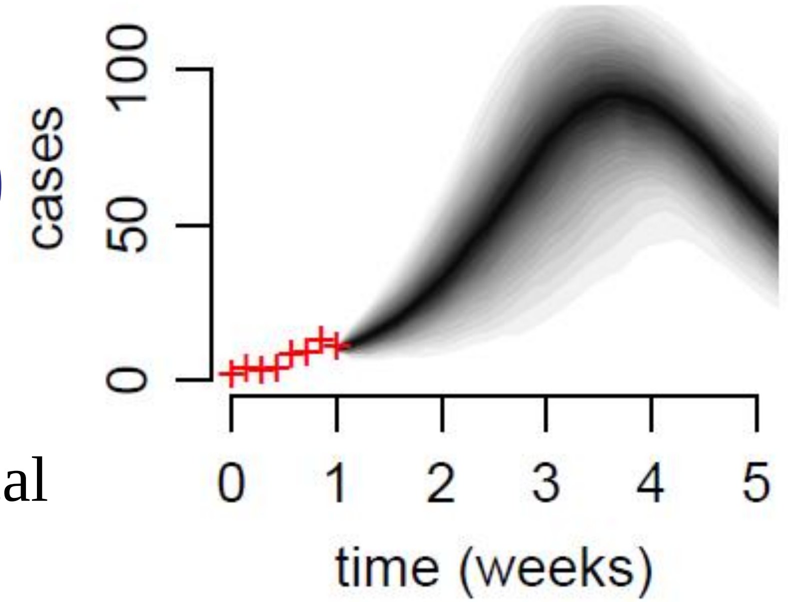


Figure: Alex Cook (NUS)

1. Parametric uncertainty

Parameter estimates given empirical data of limited sample size

2. Stochasticity

Chance structure given stochasticity

3. Serial dependence

Today's cases are caused by those in the past

Bayesian inference + SIR model

=> Permits these at once

Frequent public health questions for real time forecasting

1. How bad so far?

   => How many cases will there be in the next week?

   => How serious the total outbreak size will be?

2. Has it peaked? How high is the peak?

3. Has it been over? When will it be over?

4. What interventions to be implemented?

Empirical data: Influenza

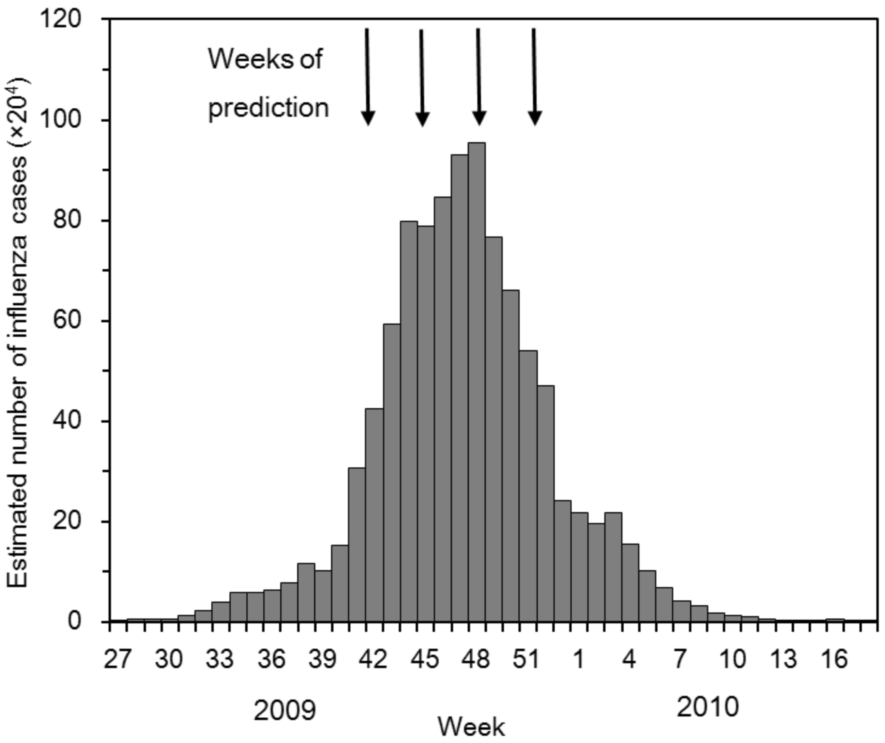
Infectious Disease Control Law

Reported by week

(can be structured by prefecture)

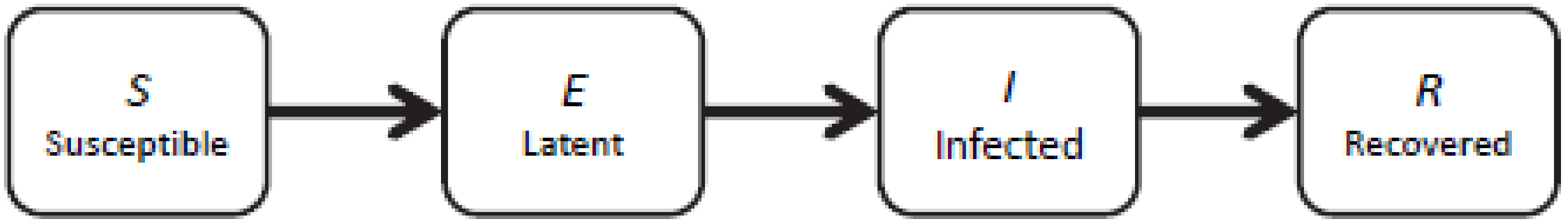
Influenza-“like” illness

Tip of iceberg (medical attendance)



Nishiura. Biomed Eng Online (2011)

Simulation model



$$S_t = S_{t-1} - A_t, \qquad A_t \sim \mathrm{Bin}\,(S_{t-1}, 1-\exp\{-[\epsilon+\beta I_{t-1}]/N\})$$

$$E_t = E_{t-1} + A_t - B_t, \qquad B_t \sim \mathrm{Bin}\,(E_{t-1}, 1-\exp\{-\lambda\})$$

$$I_t = I_{t-1} + B_t - C_t, \qquad C_t \sim \mathrm{Bin}\,(I_{t-1}, 1-\exp\{-\gamma\})$$

$$R_t = R_{t-1} + C_t$$

(Total population $N = S_t + E_t + I_t + R_t$)

letent period λ^{-1}, infectious period γ^{-1}, transmissibility β, baseline ILI ratio ϵ

# of latent A_t, infected B_t, and recovered C_t people in week t

$$k \sim \mathrm{Bin}\,(n, p) \quad \Leftrightarrow \quad \Pr[k] = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$ (binomial distribution)

Slide: Modified Masa Saito’s slide

- Observation model – connects simulation and data

$$D_t \sim \mathrm{Pois}\left[N_t \delta_{t\,(\mathrm{mod}\ 7)} \left(\phi + \frac{0.83 \times I_t}{1730} \right) \right]$$

Reported cases at week *t* (Data) D_t

# of GPFDs replied valid reponces N_t (≤ 23)

GPFD attendance ratio $= 0.83$, # of GPFDs =1730

$$k \sim \mathrm{Pois}\,(\lambda) \quad \Leftrightarrow \quad \Pr[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Slide: Modified Masa Saito's slide

Particle filter (nothing special)

Provides iteratively Monte Carlo approximation
of filtered and predictive distributions

$$p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t-1}) \simeq X_{t|t-1} \equiv \left[\boldsymbol{x}_{t|t-1}^{(1)}, \boldsymbol{x}_{t|t-1}^{(2)}, \ldots, \boldsymbol{x}_{t|t-1}^{(M)} \right]$$ Predictive

$$p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}) \simeq X_{t|t} \equiv \left[\boldsymbol{x}_{t|t}^{(1)}, \boldsymbol{x}_{t|t}^{(2)}, \ldots, \boldsymbol{x}_{t|t}^{(M)} \right]$$ Filtered

where observation vars $\boldsymbol{y}_n$ and latent vars $\boldsymbol{x}_n$ follow

$$x_n \sim f\left(\bullet \mid x_{n-1}, \theta\right)$$ "look only one step behind"

$$y_n \sim g\left(\bullet \mid x_n\right)$$ "look only the present"

for some distributions $f(x_n \mid x_{n-1}, \theta)$ and $g(y_n \mid x_n)$.

Slide: Modified Masa Saito's slide

State Space model

Simulation

$$
\begin{aligned}
S_t &= S_{t-1} - A_t, & A_t &\sim \mathrm{Bin}\,(S_{t-1}, 1-\exp\{-[\epsilon+\beta I_{t-1}]/N\}) \\
E_t &= E_{t-1} + A_t - B_t, & B_t &\sim \mathrm{Bin}\,(E_{t-1}, 1-\exp\{-\lambda\}) \\
I_t &= I_{t-1} + B_t - C_t, & C_t &\sim \mathrm{Bin}\,(I_{t-1}, 1-\exp\{-\gamma\}) \\
R_t &= R_{t-1} + C_t
\end{aligned}
$$

Observation

$$D_t \sim \mathrm{Pois}\left[N_t \delta_{t\,(\mathrm{mod}\,7)}\left(\phi + \frac{0.83 \times I_t}{1730}\right)\right]$$

Simulation

$$x_n \sim f(\bullet \mid x_{n-1}, \theta)$$

Observation

$$D_n \sim g(\bullet \mid x_n)$$

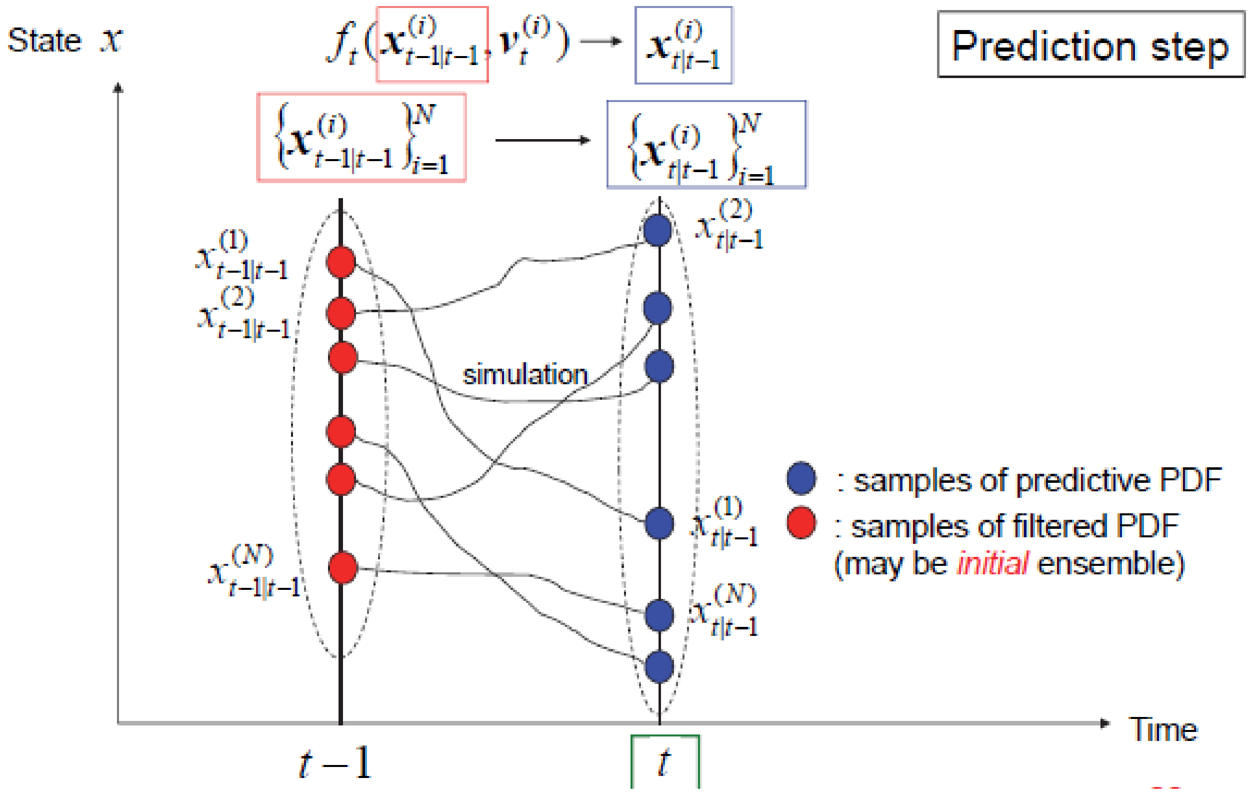
Bayesian prediction

$$\Pr\,(x_t \mid D_s, \cdots, D_1)$$

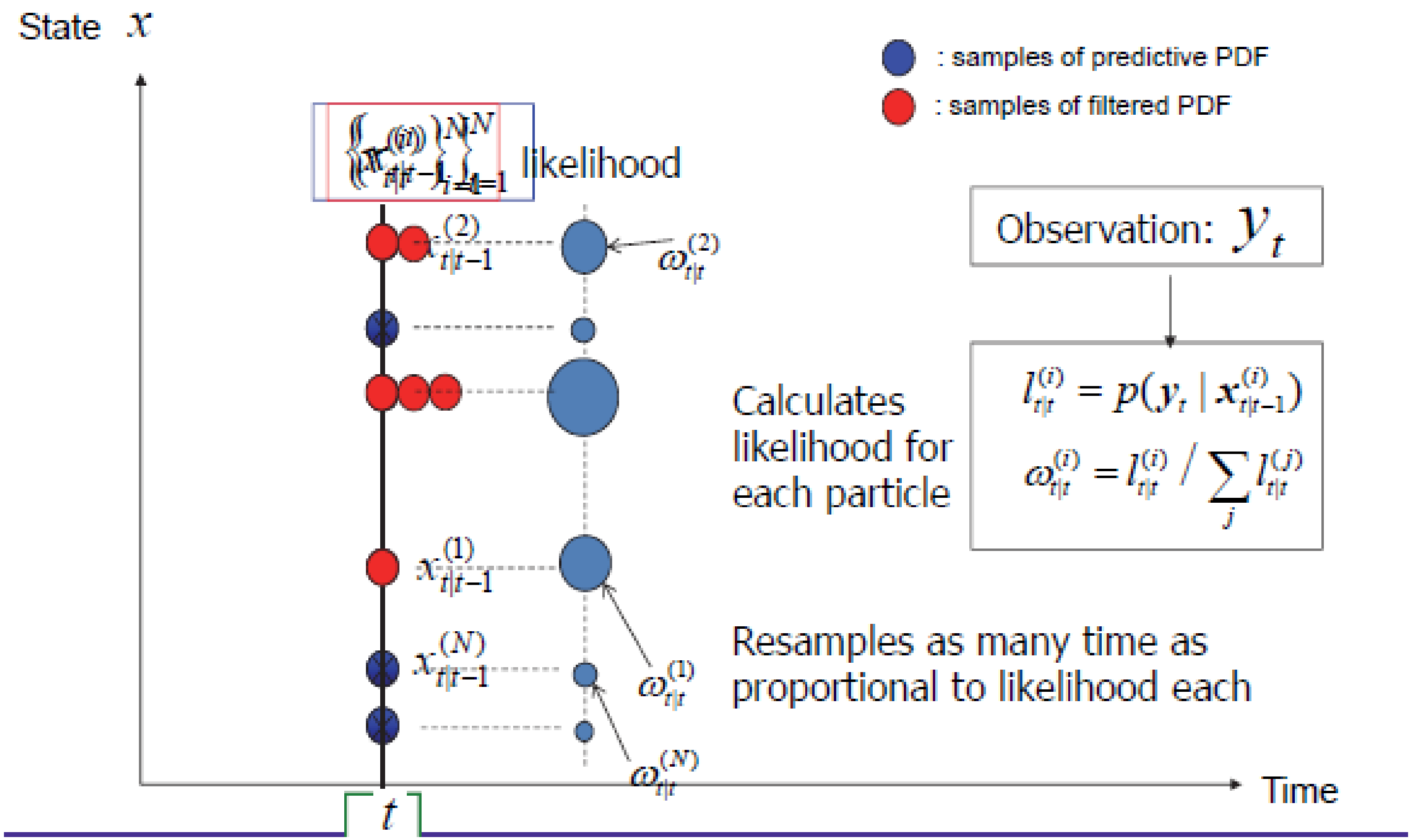
Particle filter algorithm => Monte Carlo estimation

$$\widehat{\Pr}\,(x_t \mid D_s, \cdots, D_1) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(x_t - x_t^{(i)})$$

Particle Filter: Prediction Step



1. Swarm of *particles*
2. Particles associated with simulations

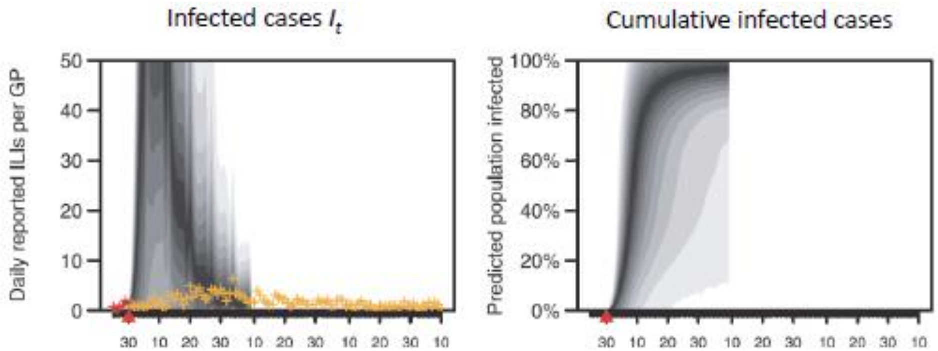


3 Simulations weighted by likelihood data conditional on particle

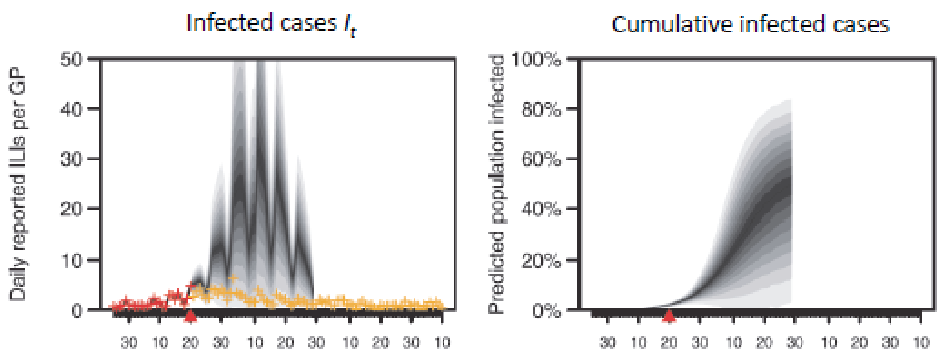
4 $\longrightarrow$ weighted sample

- Computationally light
- Predictions facile

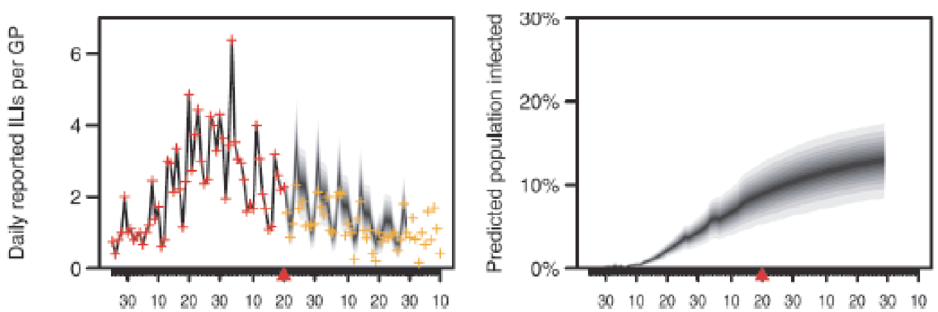
June 30, 2009



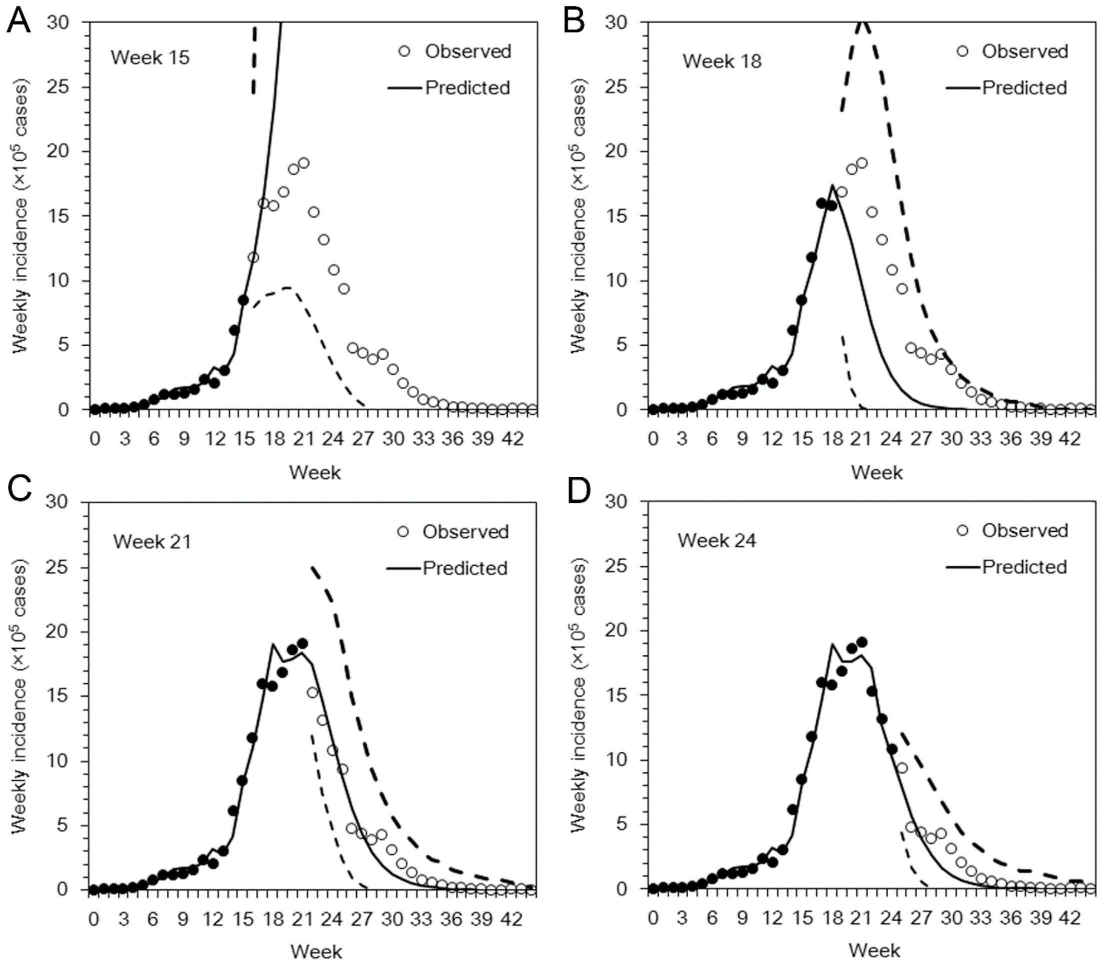
July 20, 2009



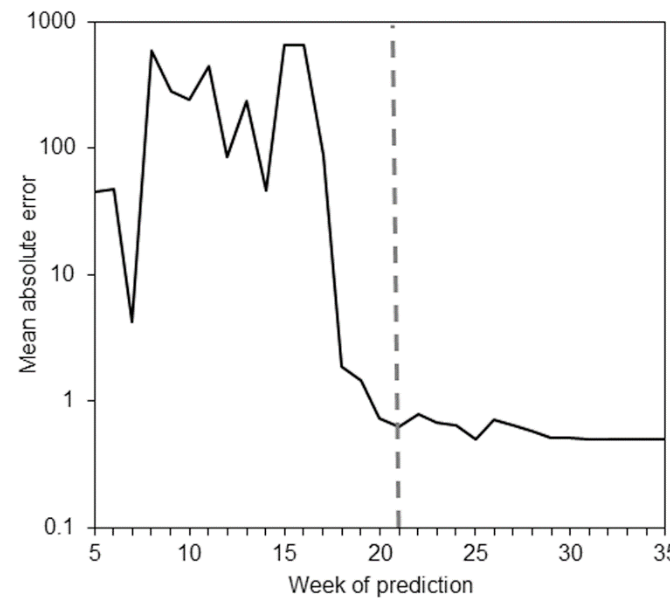
August 20, 2009



Ong et al. PLoS One 2011

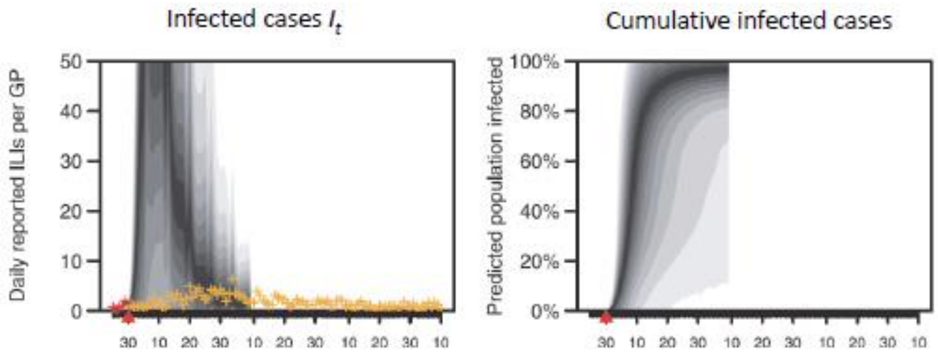


- Huge credible interval before the peak
- Due to exponential growth nature of the cases



Nishiura. Biomed Eng Online (2011)

Changing trend of ID forecasting



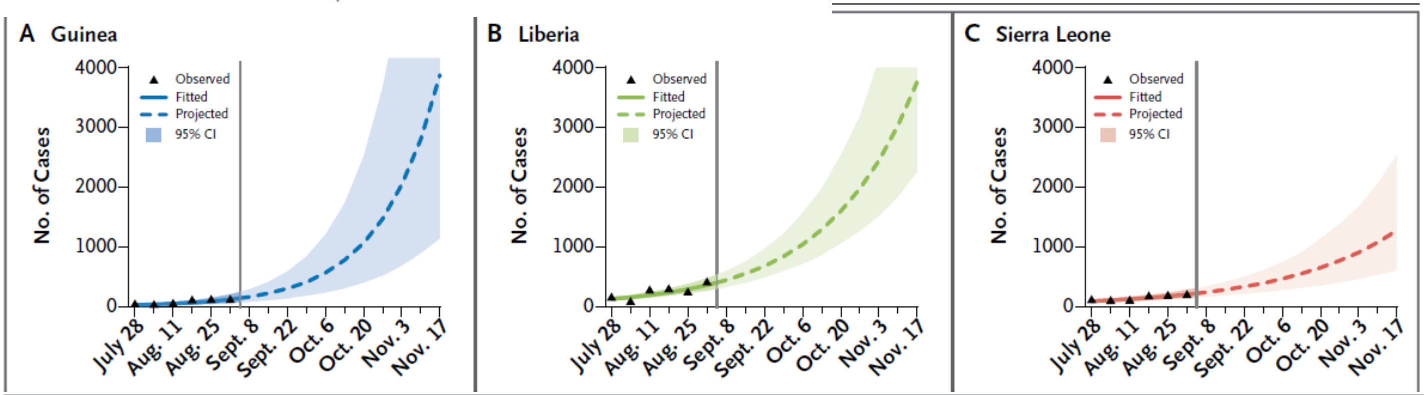
1. Not entire epidemic curve => Real time update of shorter forecast (validity using the near future data)
2. Real time advice for interventions
3. Improved predictive performance: Geography
4. Improved… (2): Climatological data
5. Improved… (3): Genome

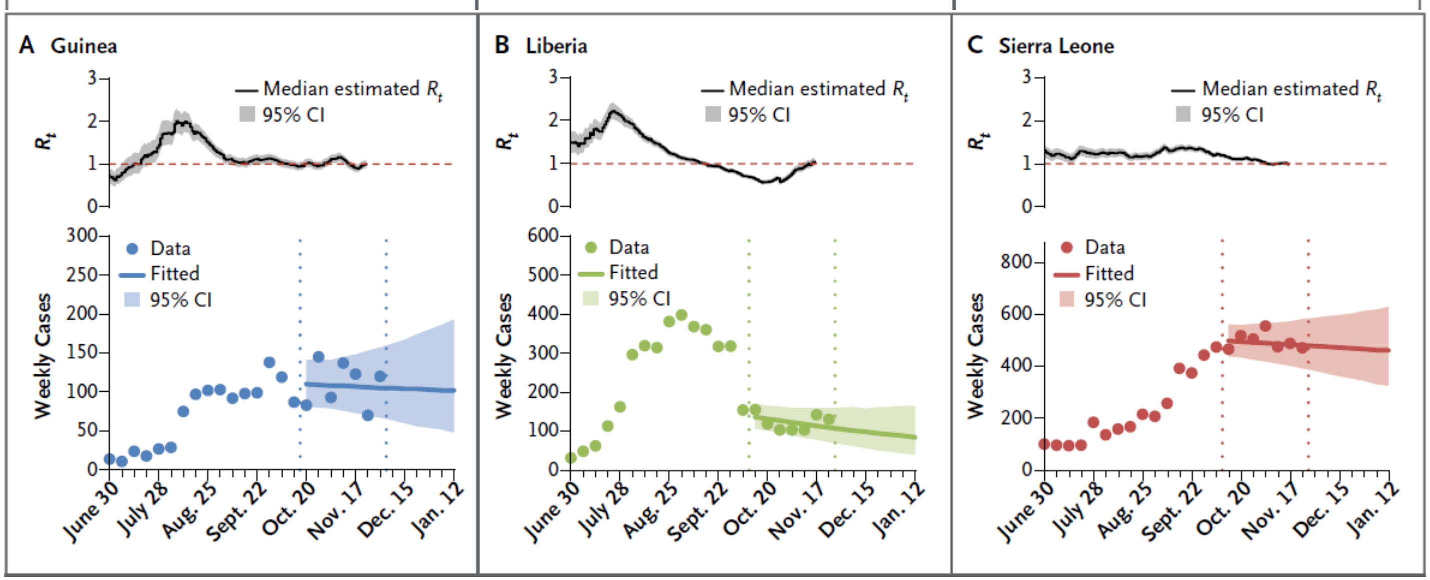
ORIGINAL ARTICLE

Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections

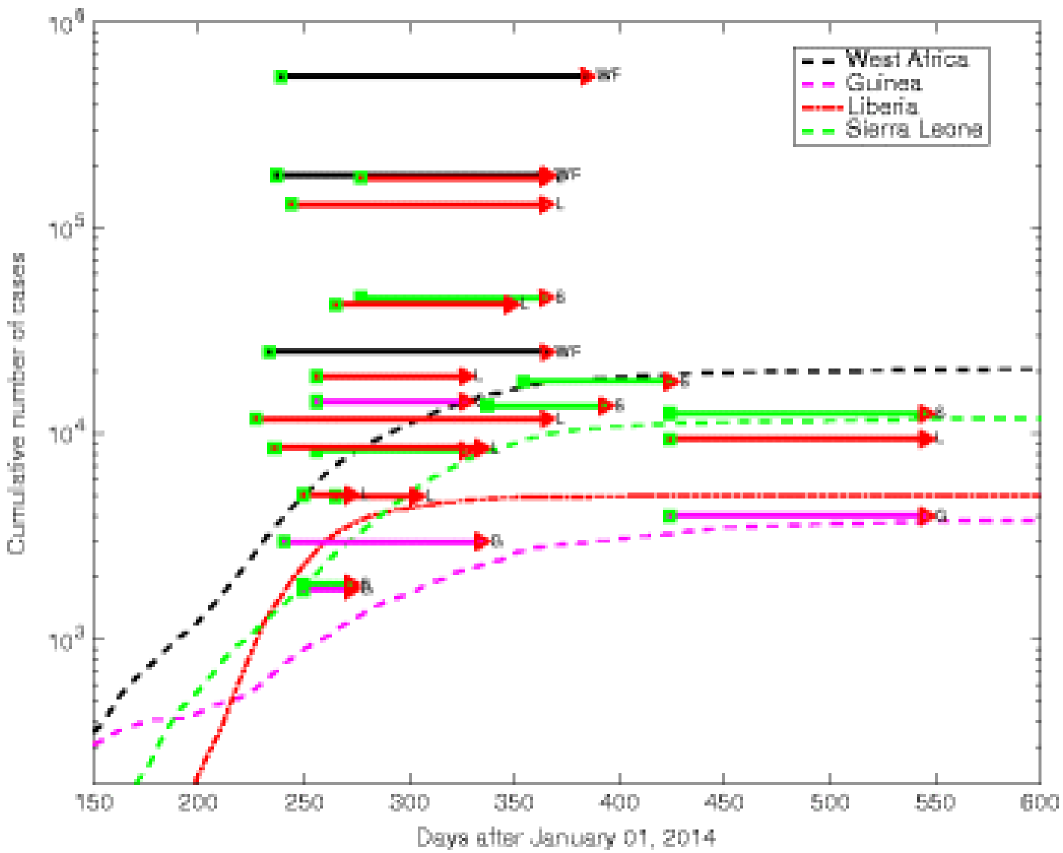
WHO Ebola Response Team*

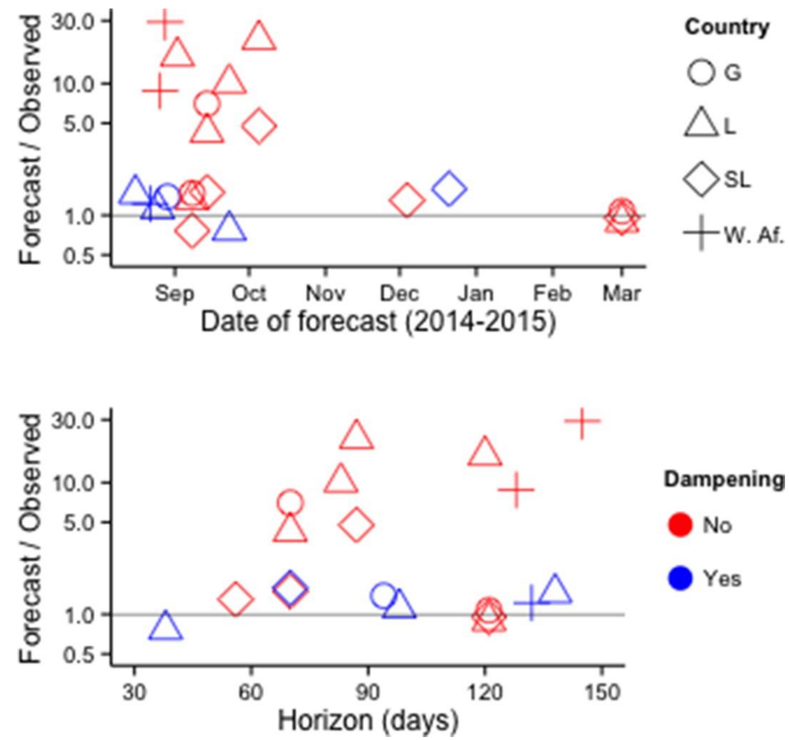
Ebola virus disease





WHOERT. N Engl J Med 2014;371:1481-95

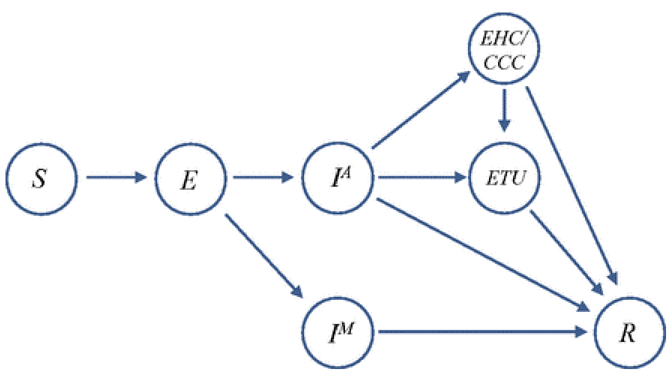


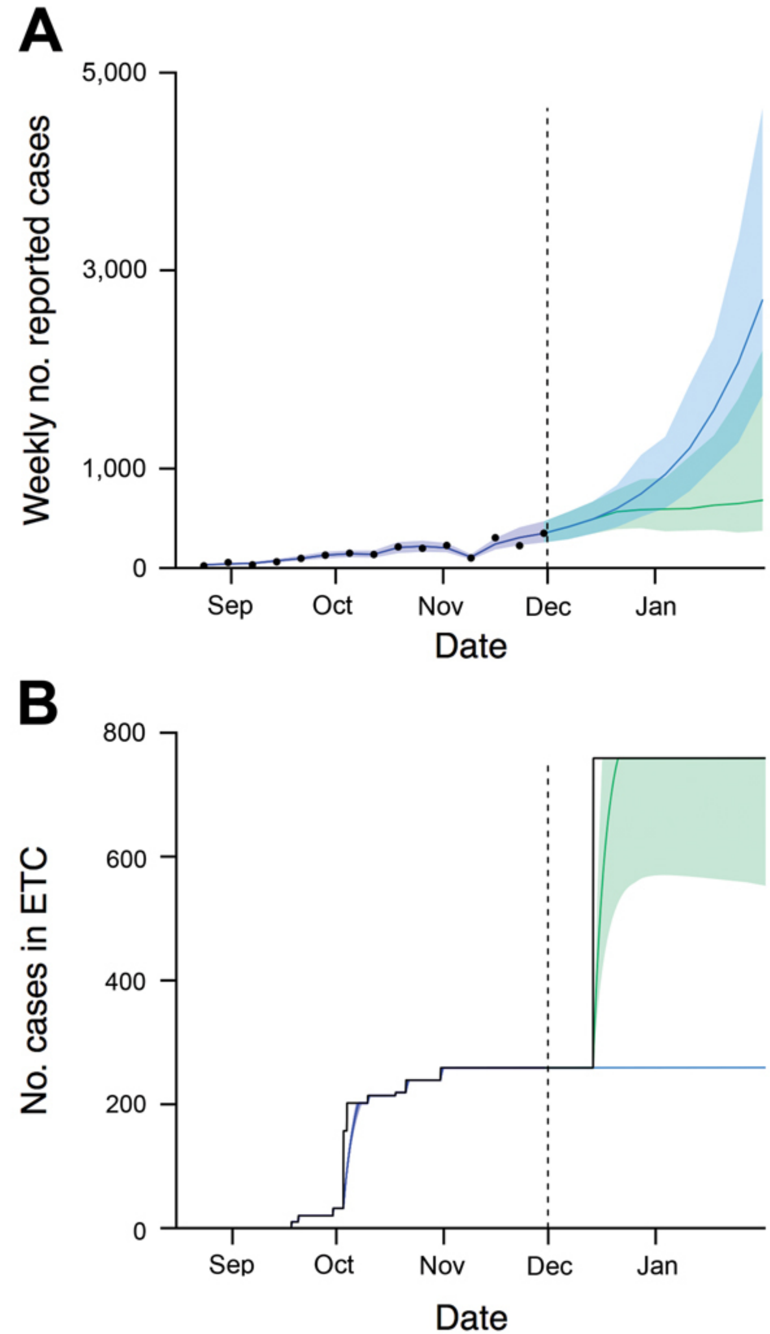


Chowell et al., BMC Medicine2017;15:42

Chretien, et al., eLife 2015;4:e09186

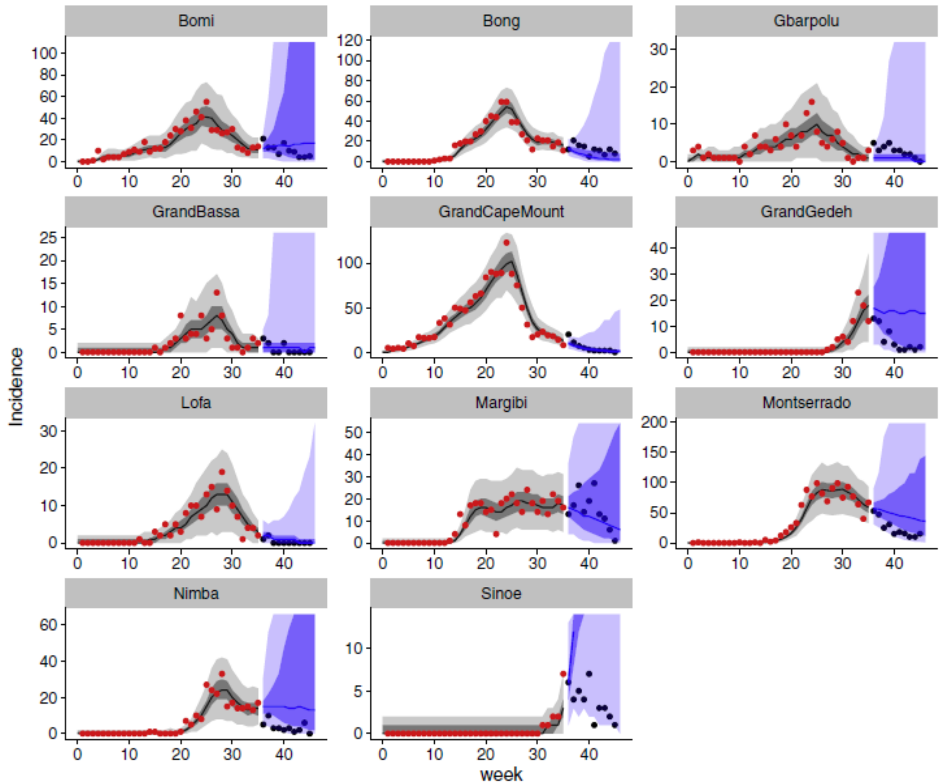
Advised interventions: Hospitals! Human resources!

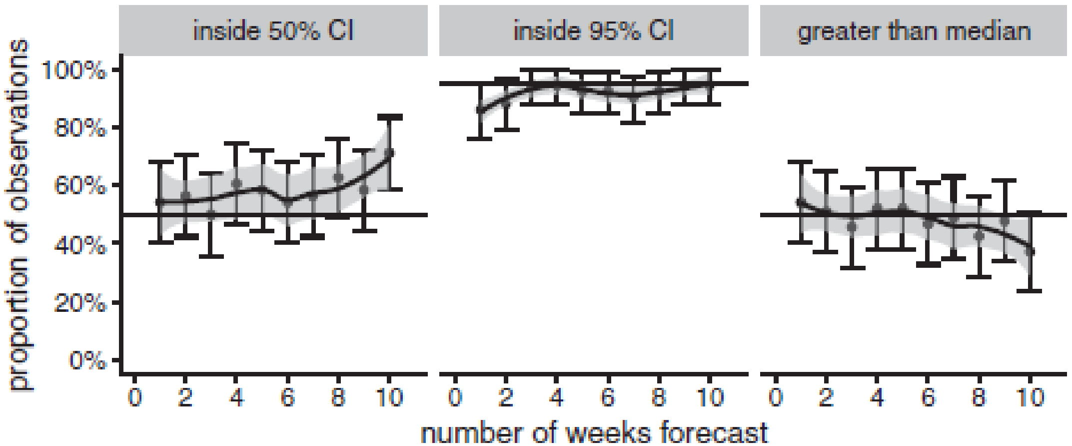




Kucharski et al., PNAS 2015:112:114366-71

Spatiotemporal model : Reduced uncertainty





Funk, et al., Epidemics 2016

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

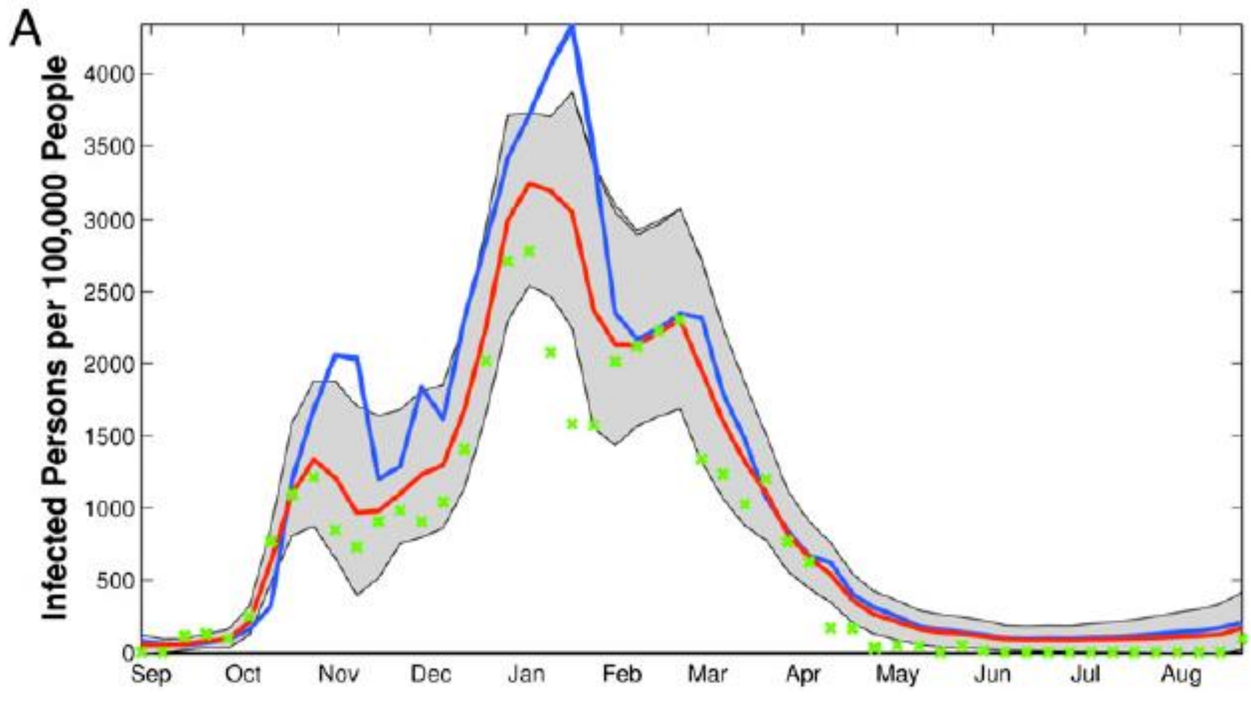
$$\frac{dI}{dt} = \frac{\beta(t)IS}{N} - \frac{I}{D} + \alpha,$$

$$R_0(t) = \beta(t)D.$$

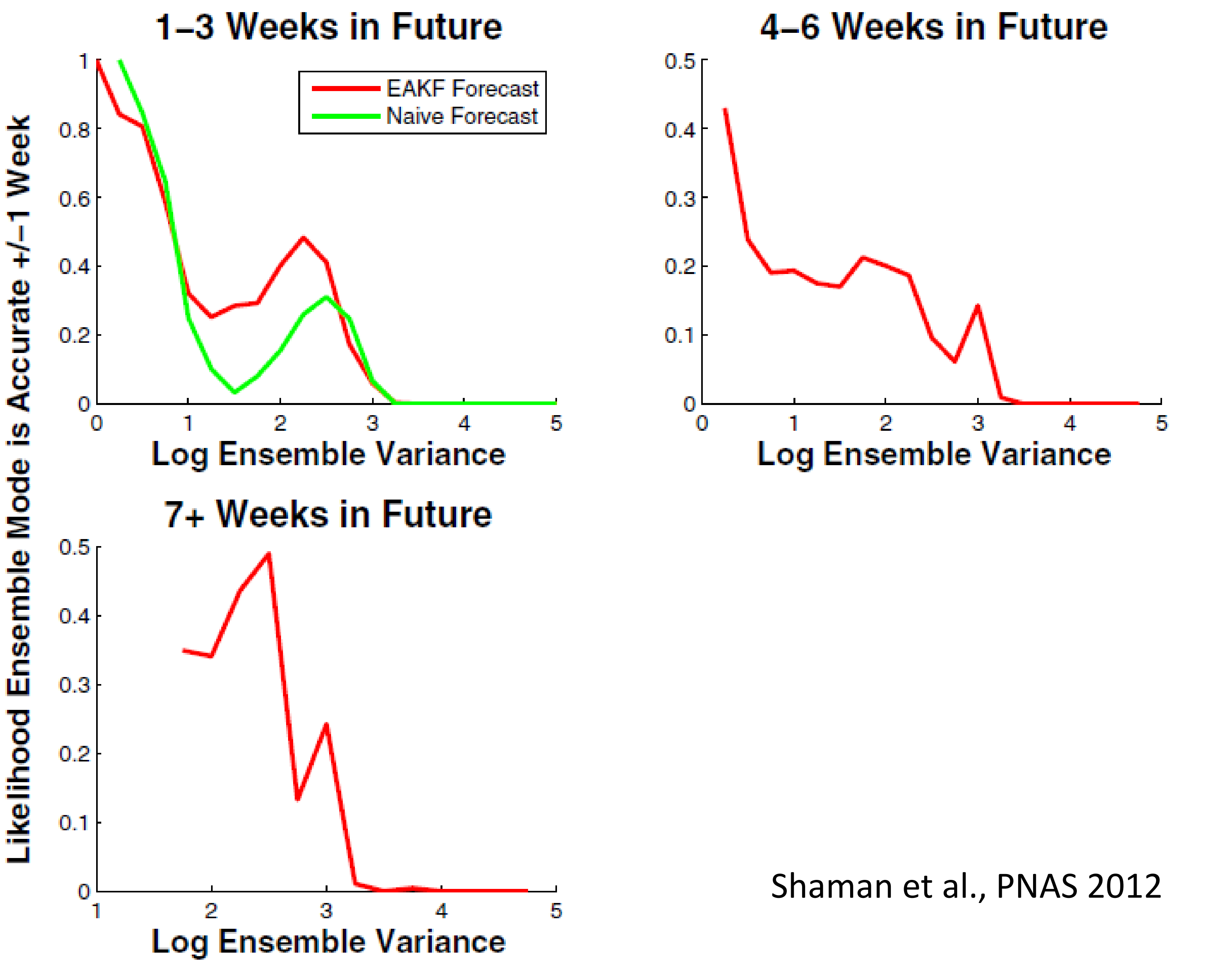
$$R_0(t) = \exp(a \times q(t) + b) + R_{0\min},$$

$q(t)$: humidity

Plug in humidity onto epi model (Case study of USA)



Shaman et al., PNAS 2012



Shaman et al., PNAS 2012

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics

Wan Yang[1]*, Alicia Karspeck[2], Jeffrey Shaman[1]

Three particle filters—

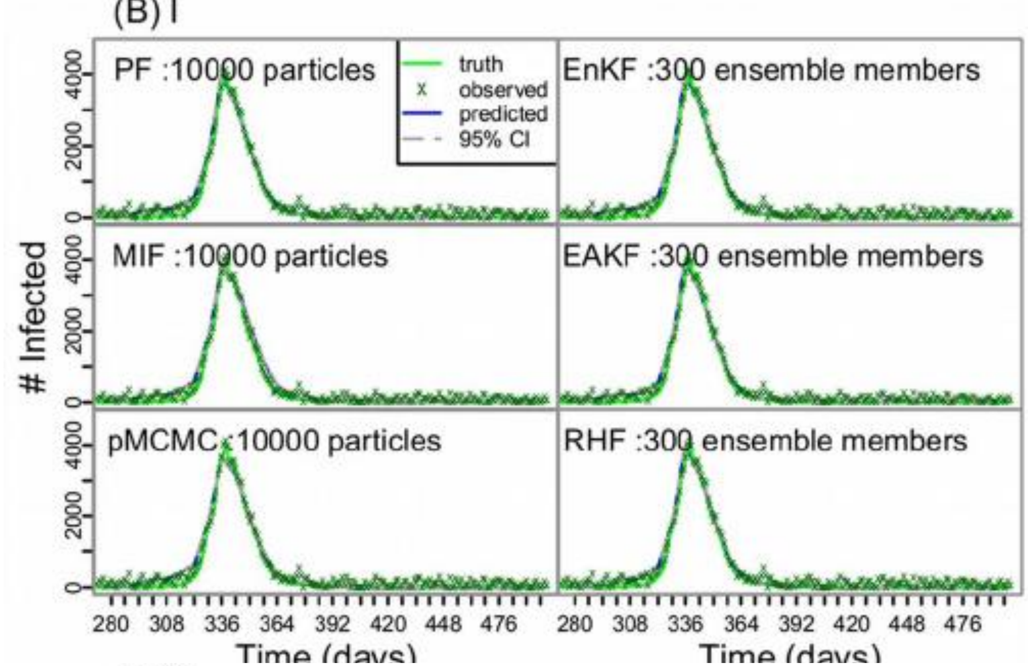
a basic particle filter (PF) with resampling and regularization,
maximum likelihood estimation via iterated filtering (MIF),
and particle Markov chain Monte Carlo (pMCMC)

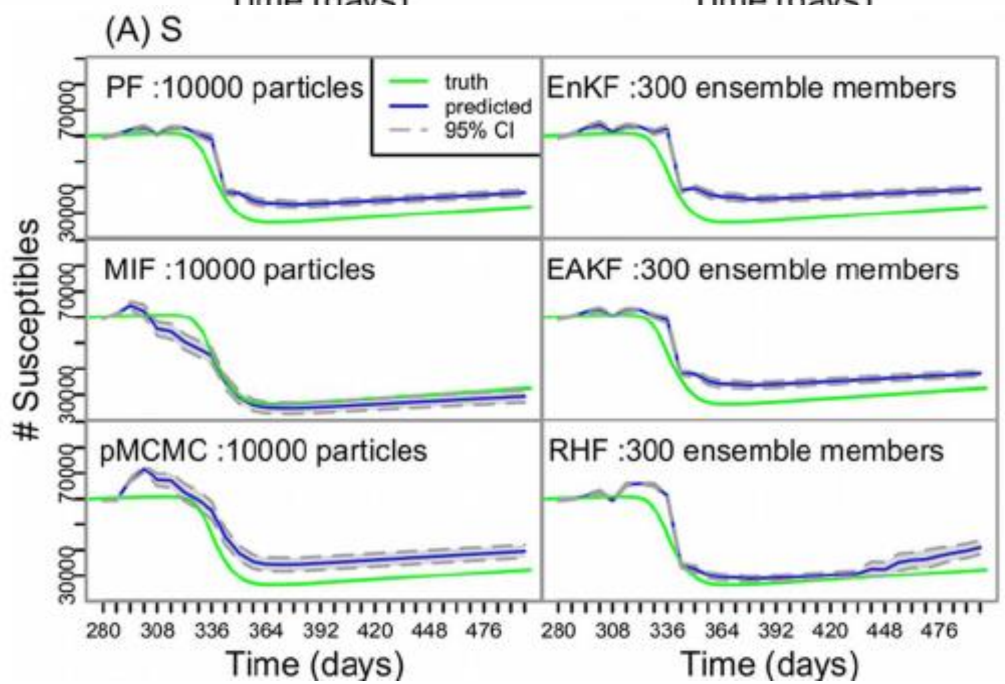
and three ensemble filters—

the ensemble Kalman filter (EnKF),
the ensemble adjustment Kalman filter (EAKF),
and the rank histogram filter (RHF)

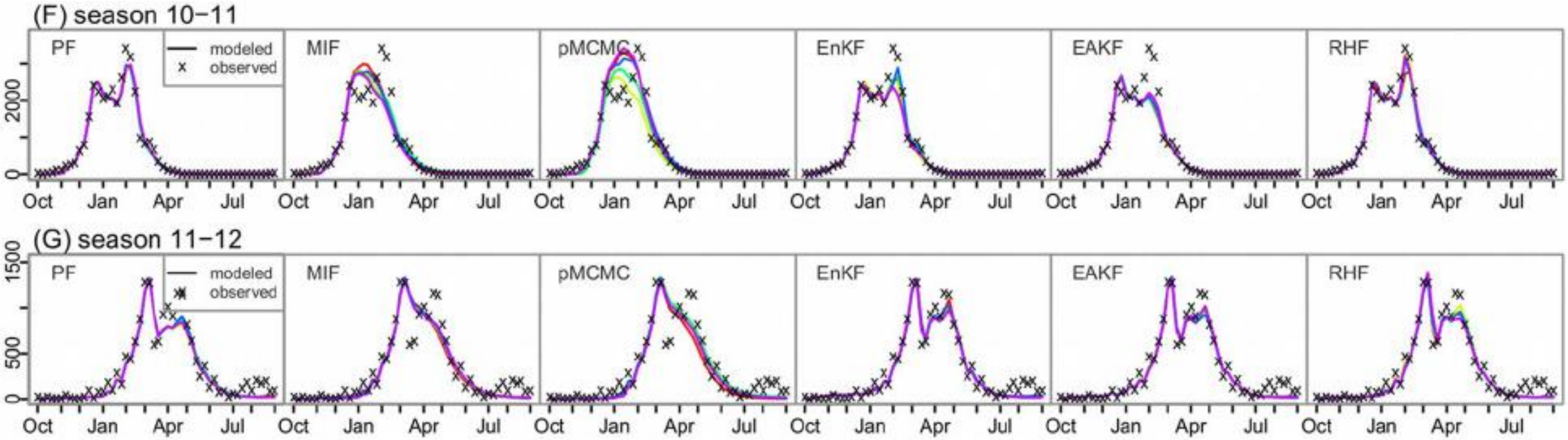
retrospectively forecast the historical incidence time series of seven influenza epidemics during 2003–2012, for 115 cities in the United States.

Yang W, et al., 2014 PLoS Comp Biol





Yang W, et al., 2014 PLoS Comp Biol



3 findings:

basic PF are more capable of faithfully recreating historical influenza incidence time series, while the MIF and pMCMC do not perform as well for multimodal outbreaks.
--- adjust model parameters continually (at each prediction-update cycle)

For forecast of the week with the highest influenza activity, the accuracies of the six model-filter frameworks are comparable;

the ensemble filters are more accurate predicting peaks in the past.

Yang W, et al., 2014 PLoS Comp Biol

Summary

1. PF and its relatives (e.g. pMCMC) are increasingly used for epidemic forecasting
2. Single curve prediction far before the peak yields huge credible interval
3. Spatial data, climatological data=> successful in reducing uncertainty
4. Interventions: answered in real time during Ebola epidemic

Japan: far less advanced... Academically attacked when an epidemic happens in the country