Applications of Likelihood-based inference with non-mechanistic and mechanistic models in infectious disease modeling

Daihai HE

Department of Applied Mathematics
Hong Kong Polytechnic University
daihai.he@polyu.edu.hk

Webinar of International Society of Disease Surveillance

June 21, 2019

Outline

Outline

Likelihood-based inference

- The likelihood function is the joint probability distribution of observed data expressed as a function of statistical parameters.
- The likelihood function describes a surface over the domain of permissible parameter values.
- The value that is most likely to be the parameter of the joint probability distribution underlying the observed data is the peak of the surface.
- The procedure for obtaining these arguments of the maximum of the likelihood function is known as maximum likelihood estimation.

`https://en.wikipedia.org/wiki/Likelihood_function`

Iterated Filtering (Ionides, E. L., Breto, C. and King, A. A. (2006))

- Iterated filtering algorithms are a tool for maximum likelihood inference on partially observed dynamical systems.
- Stochastic perturbations to the unknown parameters are used to explore the parameter space. Applying sequential Monte Carlo (the particle filter) to this extended model results in the selection of the parameter values that are more consistent with the data.
- Appropriately constructed procedures, iterating with successively diminished perturbations, converge to the maximum likelihood estimate.
- Iterated filtering methods have so far been used most extensively to study infectious disease transmission dynamics.

`https://en.wikipedia.org/wiki/Iterated_filtering`

Iterated Filtering (Ionides, E. L., Breto, C. and King, A. A. (2006))

- The data are a time series $y_1, \ldots, y_N$ collected at times $t_1 < t_2 < \cdots < t_N$.
- The dynamic system is modeled by a Markov process $X(t)$ which is generated by a function $f(x, s, t, \theta, W)$ in the sense that $X(t_n) = f(X(t_{n-1}), t_{n-1}, t_n, \theta, W)$ where θ is a vector of unknown parameters and W is some random quantity that is drawn independently each time $f(.)$ is evaluated.
- An initial condition $X(t_0)$ at some time $t_0 < t_1$ is specified by an initialization function, $X(t_0) = h(\theta)$.
- A measurement density $g(y_n|X_n, t_n, \theta)$ completes the specification of a partially observed Markov process (POMP).

`https://en.wikipedia.org/wiki/Iterated_filtering`

Iterated Filtering (Ionides, E. L., Breto, C. and King, A. A. (2006))

Input: A partially observed Markov model specified as above; Monte Carlo sample size J; number of iterations M; cooling parameters $0 < a < 1$ and b; covariance matrix Φ; initial parameter vector $\theta^{(1)}$

for $m = 1$ to M

draw $\Theta_F(t_0, j) \sim \text{Normal}(\theta^{(m)}, ba^{m-1}\Phi)$ for $j = 1, \ldots, J$

set $X_F(t_0, j) = h\big(\Theta_F(t_0, j)\big)$ for $j = 1, \ldots, J$

set $\bar{\theta}(t_0) = \theta^{(m)}$

for $n = 1$ to N

draw $\Theta_P(t_n, j) \sim \text{Normal}(\Theta_F(t_{n-1}, j), a^{m-1}\Phi)$ for $j = 1, \ldots, J$

set $X_P(t_n, j) = f(X_F(t_{n-1}, j), t_{n-1}, t_n, \Theta_P(t_n, j), W)$ for $j = 1, \ldots, J$

set $w(n, j) = g(y_n | X_P(t_n, j), t_n, \Theta_P(t_n, j))$ for $j = 1, \ldots, J$

draw $k_1, \ldots, k_J$ such that $P(k_j = i) = w(n, i) / \sum_\ell w(n, \ell)$

set $X_F(t_n, j) = X_P(t_n, k_j)$ and $\Theta_F(t_n, j) = \Theta_P(t_n, k_j)$ for $j = 1, \ldots, J$

set $\bar{\theta}_i(t_n)$ to the sample mean of $\{\Theta_{F,i}(t_n, j), j = 1, \ldots, J\}$, where the vector Θ_F has components $\{\Theta_{F,i}\}$

set $V_i(t_n)$ to the sample variance of $\{\Theta_{P,i}(t_n, j), j = 1, \ldots, J\}$

set $\theta_i^{(m+1)} = \theta_i^{(m)} + V_i(t_1) \sum_{n=1}^{N} V_i^{-1}(t_n)(\bar{\theta}_i(t_n) - \bar{\theta}_i(t_{n-1}))$

Output: Maximum likelihood estimate $\hat{\theta} = \theta^{(M+1)}$

Plug-and-Play Inference Framework

- The outbreak is modeled as a Partially Observed Markov process (POMP) and makes use of Iterated Filtering and plug-and-play likelihood-based inference frameworks to fit the data.
- Information Criterion quantifying the tradeoff between the goodness-of-fit of a model and its complexity, is employed for model comparison.
- The simulations were conducted deploying the Euler-multinomial integration method with the time-step fixed to be one day.
- The weekly observed cases, C_i, are assumed to follow a Negative-Binomial (NB) distribution as

$$C_i \sim \text{NB}\left(n = \frac{1}{\tau}, p = \frac{1}{1+\tau Z_{h,i}}\right) \quad \text{with} \quad \text{mean} : \mu_i = Z_{h,i} \quad (1)$$

where τ denotes an over-dispersion parameter that needs to be estimated.

Information Criterion

The small-sample-size corrected Akaike's Information Criterion (AICc) is a measurement of the trade-off between model complexity and the goodness-of-fit. The AICc is given by:

$$\mathrm{AICc} = -2l(\hat{\Theta}) + 2k + \frac{2k(k+1)}{N-k-1}, \tag{2}$$

where N is the number of data points and k is the number of free parameters.
Bayesian Information Criterion (BIC) is defined as

$$\mathrm{BIC} = -2l(\hat{\Theta}) + k \ln N \tag{3}$$

Outline

Data and Question

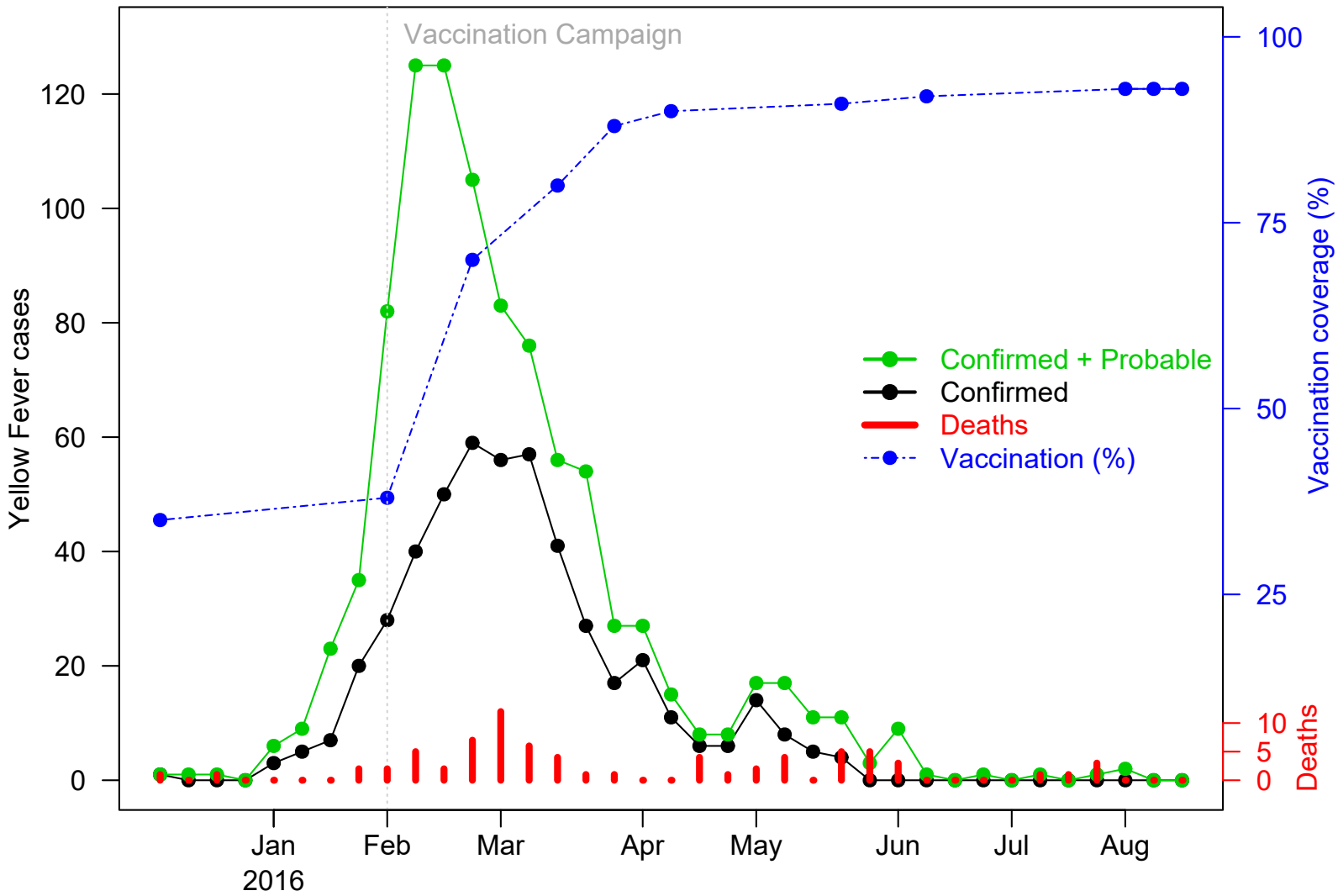


Figure: Yellow fever outbreak in Luanda from Dec 5, 2015 to Aug 18, 2016.

Yellow Fever Model

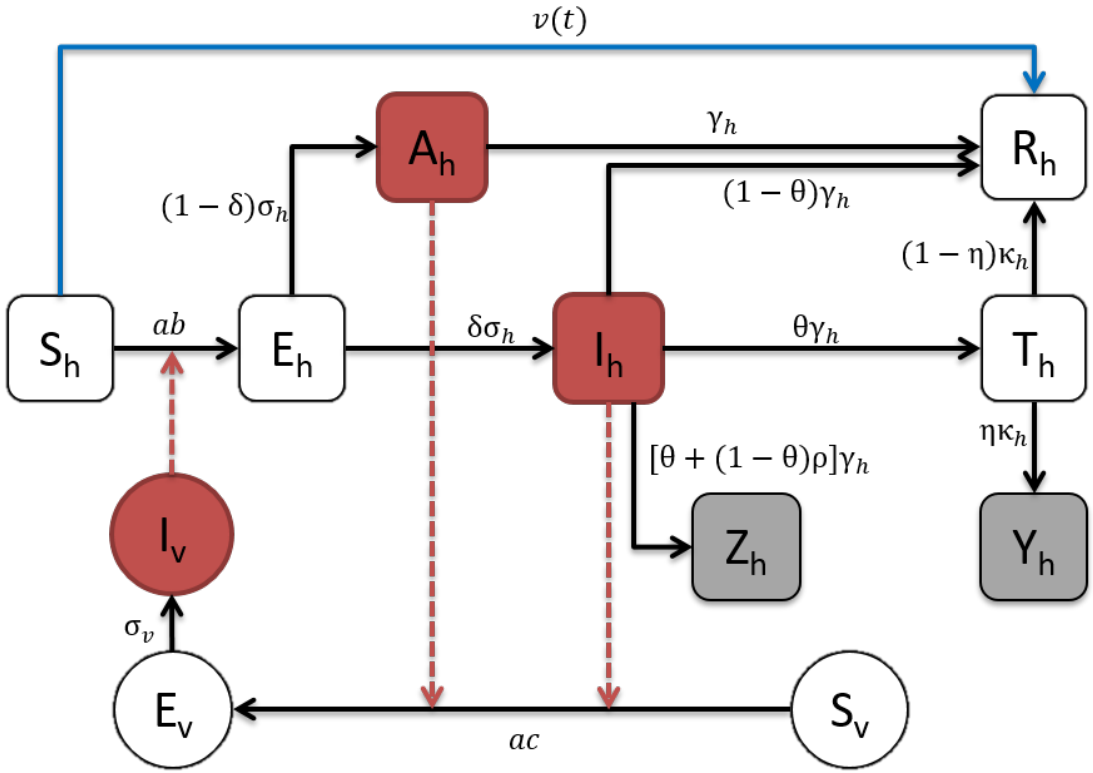


Figure: Flowchart of the yellow fever model.

Yellow Fever Model

$$
\begin{aligned}
S_h' &= -ab\frac{I_v}{N_h}S_h - v(t) && (4a)\\
E_h' &= ab\frac{I_v}{N_h}S_h - \sigma_h E_h && (4b)\\
A_h' &= (1-\delta)\cdot\sigma_h E_h - \gamma_h A_h && (4c)\\
I_h' &= \delta\sigma_h E_h - \gamma_h I_h && (4d)\\
T_h' &= \theta\gamma_h I_h - \kappa_h T_h && (4e)\\
R_h' &= v(t) + \gamma_h A_h + (1-\theta)\cdot\gamma_h I_h + (1-\eta)\cdot\kappa_h T_h && (4f)\\
S_v' &= B_v(t) - ac\frac{\psi A_h + I_h}{N_h}S_v - \mu_v S_v && (4g)\\
E_v' &= ac\frac{\psi A_h + I_h}{N_h}S_v - \sigma_v E_v - \mu_v E_v && (4h)\\
I_v' &= \sigma_v E_v - \mu_v I_v && (4i)
\end{aligned}
$$

Yellow Fever Model (non-mechanistic $m(t)$)

$$\text{weekly deaths: } Y_{h,i} = \int_{\text{week } i} \eta \kappa_h T_h \, dt \tag{5a}$$

$$\text{weekly reported cases: } Z_{h,i} = \int_{\text{week } i} [\theta + \rho \cdot (1-\theta)] \cdot \gamma_h I_h \, dt \tag{5b}$$

$$N_h = S_h + E_h + A_h + I_h + T_h + R_h + \sum_i Y_{h,i} \tag{6a}$$

$$N_v = S_v + E_v + I_v \tag{6b}$$

$$N_v(t) = m(t) \cdot N_h \tag{7}$$

where $m(t)$ is modeled as an exponential cubic spline function.

Vaccination Scenarios

A simple way to estimate the recovered (including vaccinated) population

$$R_h(t) = V(t - t_0) \cdot N_h + \varepsilon_0 + \varepsilon(t). \tag{8}$$

where $V(t)$ is the vaccination coverage, ε_0 denotes the prior immunity and $\varepsilon(t)$ denotes the cumulative infected cases up to time t.

- Baseline scenario: actual vaccination campaign as experienced in Luanda;
- Alternative scenarios: 60, 120 and 180 days ($\equiv$ no vaccination) delays of vaccination campaign

The total reported cases as well as total deaths will be evaluated by the model for each of the different vaccination scenarios.

Basic Reproduction Number

$$\mathcal{R}_0 = \sqrt{[\psi \cdot (1-\delta) + \delta] \cdot \frac{a^2 bcm}{\gamma_h} \cdot \frac{\sigma_v}{\mu_v(\sigma_v + \mu_v)}}. \tag{9}$$

or

$$\mathcal{R}_0 = [\psi \cdot (1-\delta) + \delta] \cdot \frac{a^2 bcm}{\gamma_h} \cdot \frac{\sigma_v}{\mu_v(\sigma_v + \mu_v)}. \tag{10}$$

Asymptomatic Infectivity Scenarios

1 : 85% asymptomatic ($\delta = 15\%$) and weak infectivity ($\psi = 0.1$)

2 : 85% asymptomatic ($\delta = 15\%$) and strong infectivity ($\psi = 0.5$)

Model Fitting

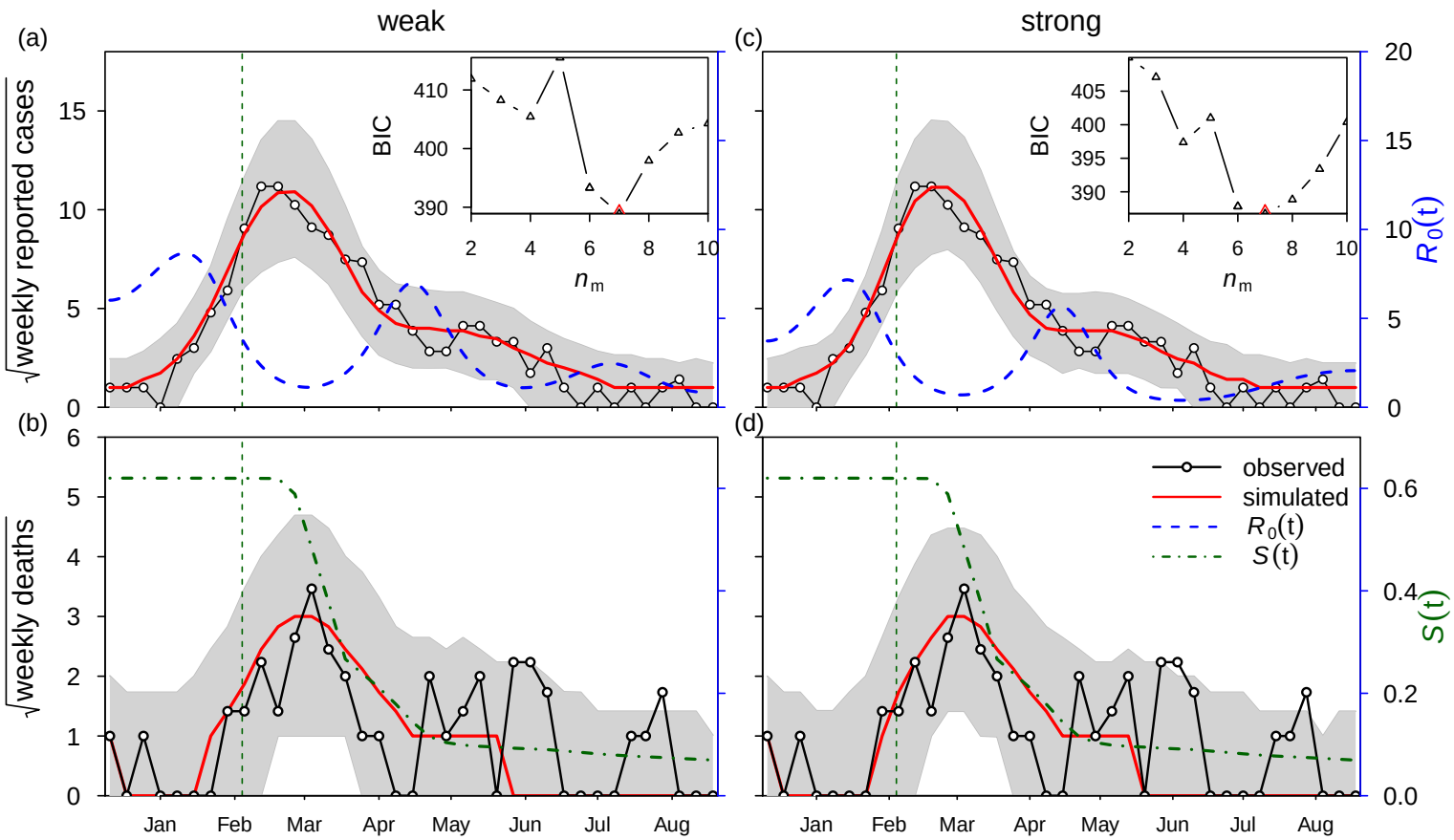


Figure: Fitting results under two scenarios of asymptomatic infectivity status.

Impact of Vaccination Delays

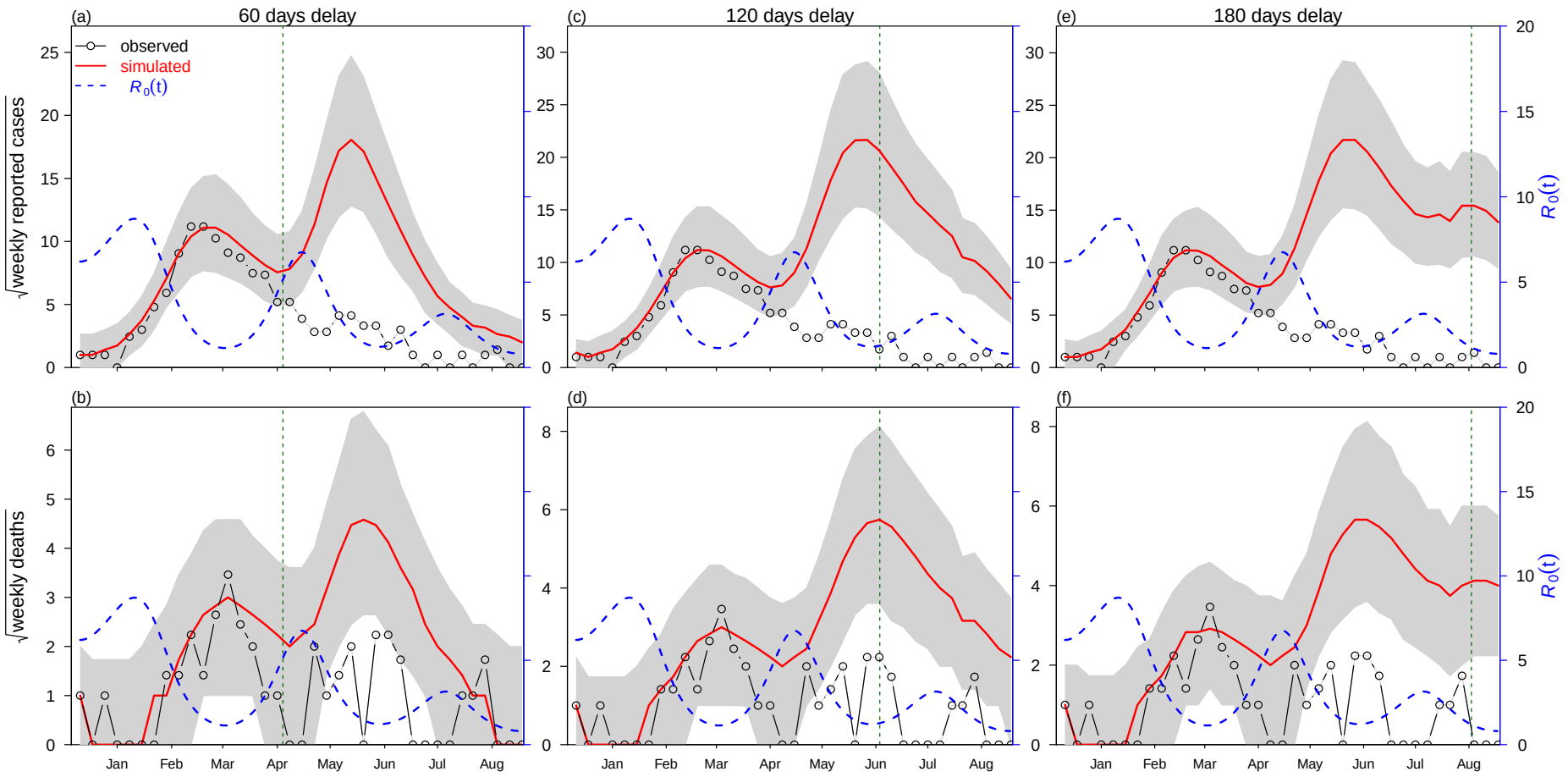


Figure: Simulation results of asymptomatic-1 scenario under three deferred vaccination campaign scenarios: 60-day delay in panels (a,b), 120-day delay in panels (c, d) and 180-day delay in panels (e,f).

Table: **Impacts of vaccination campaign delay under asymptomatic 1 scenario.**

Scenario	Total reported cases	Total deaths
Observed	941	73
Baseline model	1026 [540 , 1797]	77 [35 , 139]
60 days delay	3143 [1604 , 5584]	233 [119 , 411]
120 days delay	5450 [2751 , 9611]	400 [203 , 724]
180 days delay	6242 [3139 , 10919]	444 [226 , 787]

"$*$ $[*, *]$" denotes the simulation median with 95% Confidence Interval (C.I.).

Human behavioral model (mechanistic)

$$m(t) = m_{\text{base}} + k \cdot \exp\Big[- D_h(t - t_{\text{lag}})\Big] \tag{11}$$

Here m_{base} is a constant term, k is a parameter controlling the strength of the death-induced human reaction, $D_h(t)$ is the yellow fever deaths of time t and t_{lag} is the lag time for the population reacting to the yellow fever death situation. We use linear interpolation to convert the weekly death into a continuous time function. The fitting results for this simple human behavior model are shown in the following figure, with $t_{\text{lag}} = 1$ week fixed.

Human behavioral model

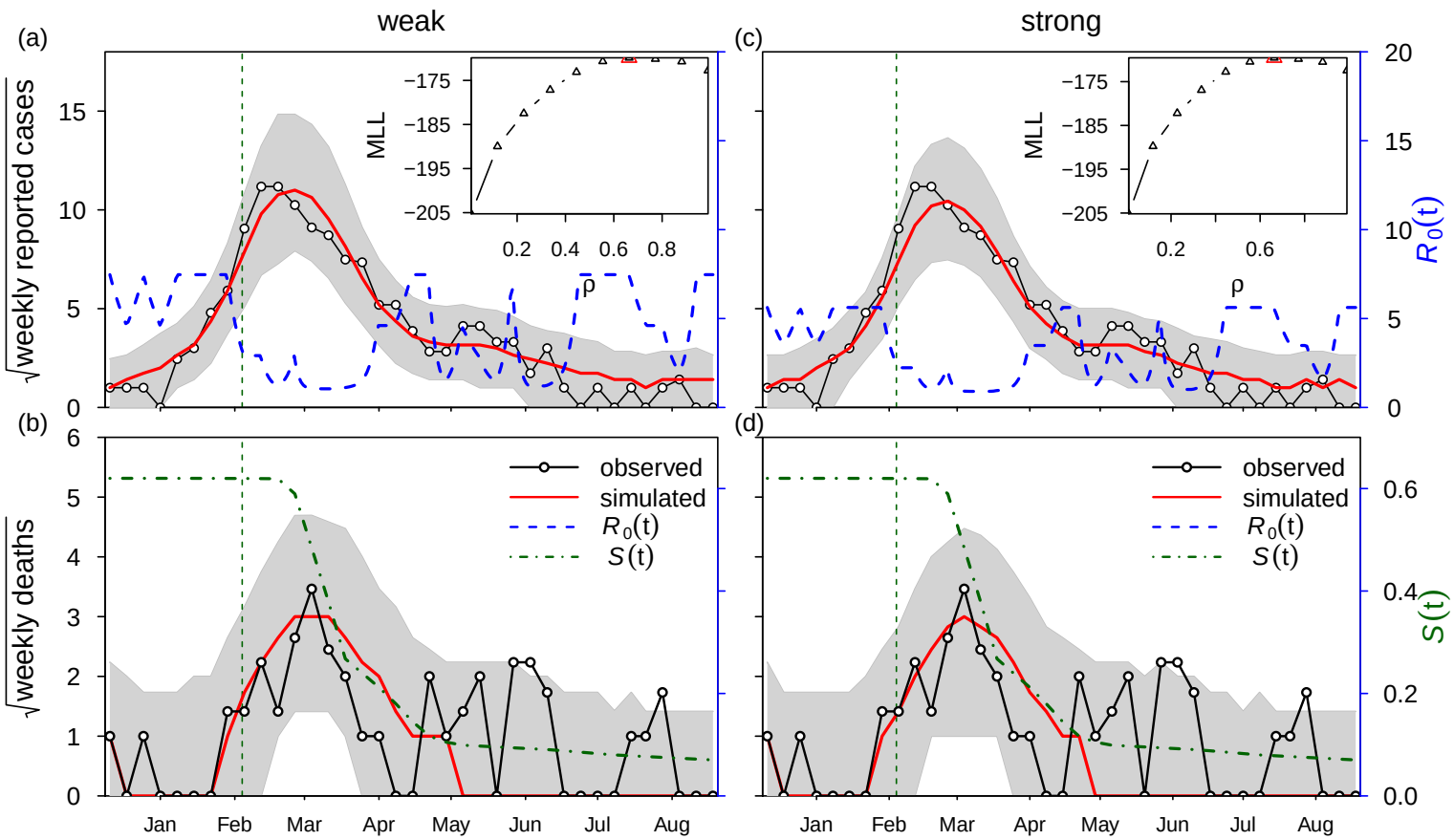
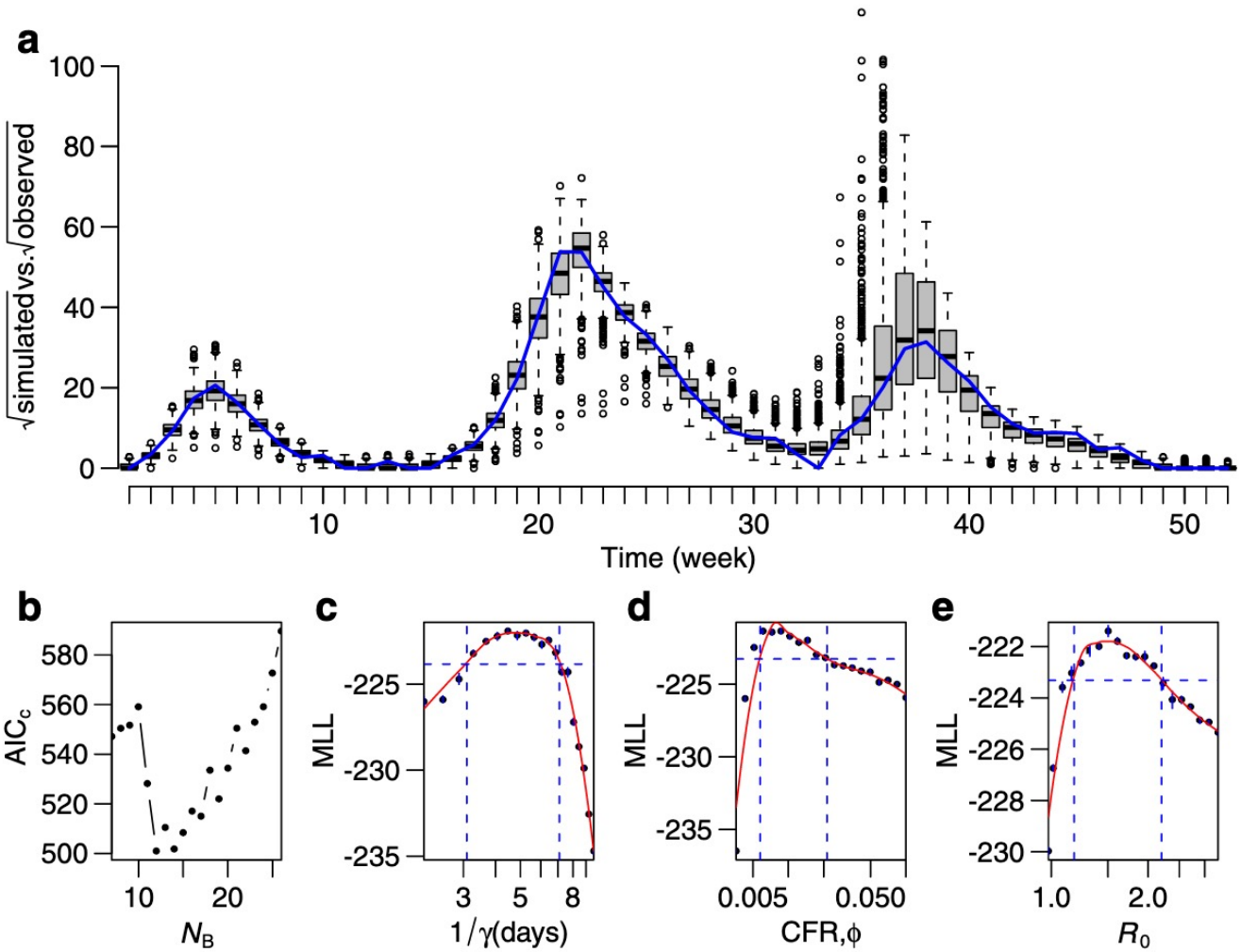


Figure: Fitting results under consideration of humans behavior (with one-week time lag, $t_{\text{lag}} = 1$).

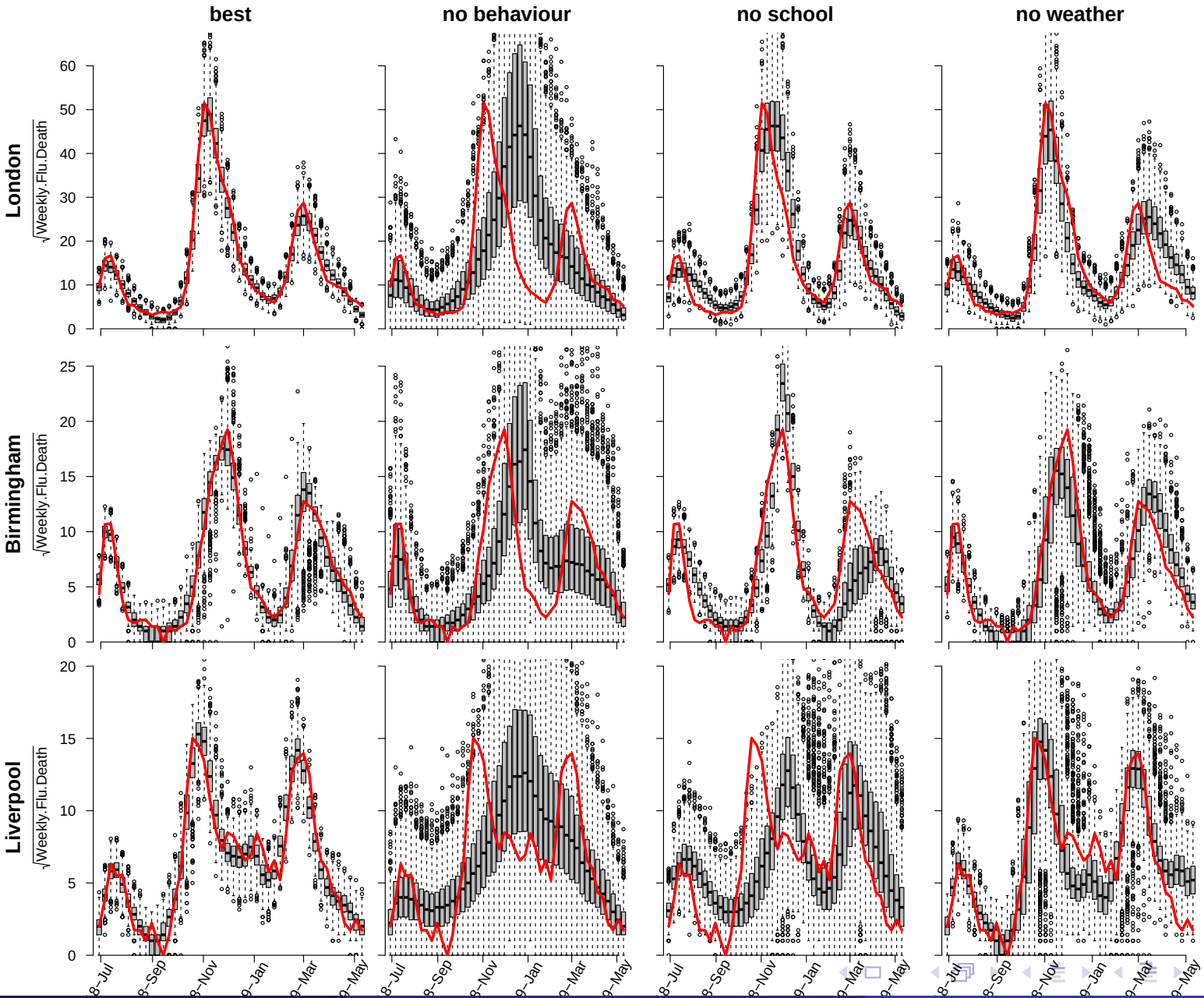
Outline

1918 Influenza in London, England (non-mechanistic model)

Fig. 2 Excess P&I mortality, realizations of the best model and likelihood profiles for four parameters. The *top panel* **a** shows the observed excess P&I mortality (*blue*, as in Fig. 1), together with box plots of 1,000 realizations of the best model. Panel **b** shows the AICc as a function of N_B, the number of B-spline basis functions. Panels **c–e** show likelihood profiles for $1/\gamma$, ϕ and $\mathcal{R}_0$ (with *horizontal lines* that define the 95% confidence intervals; see "Methods")

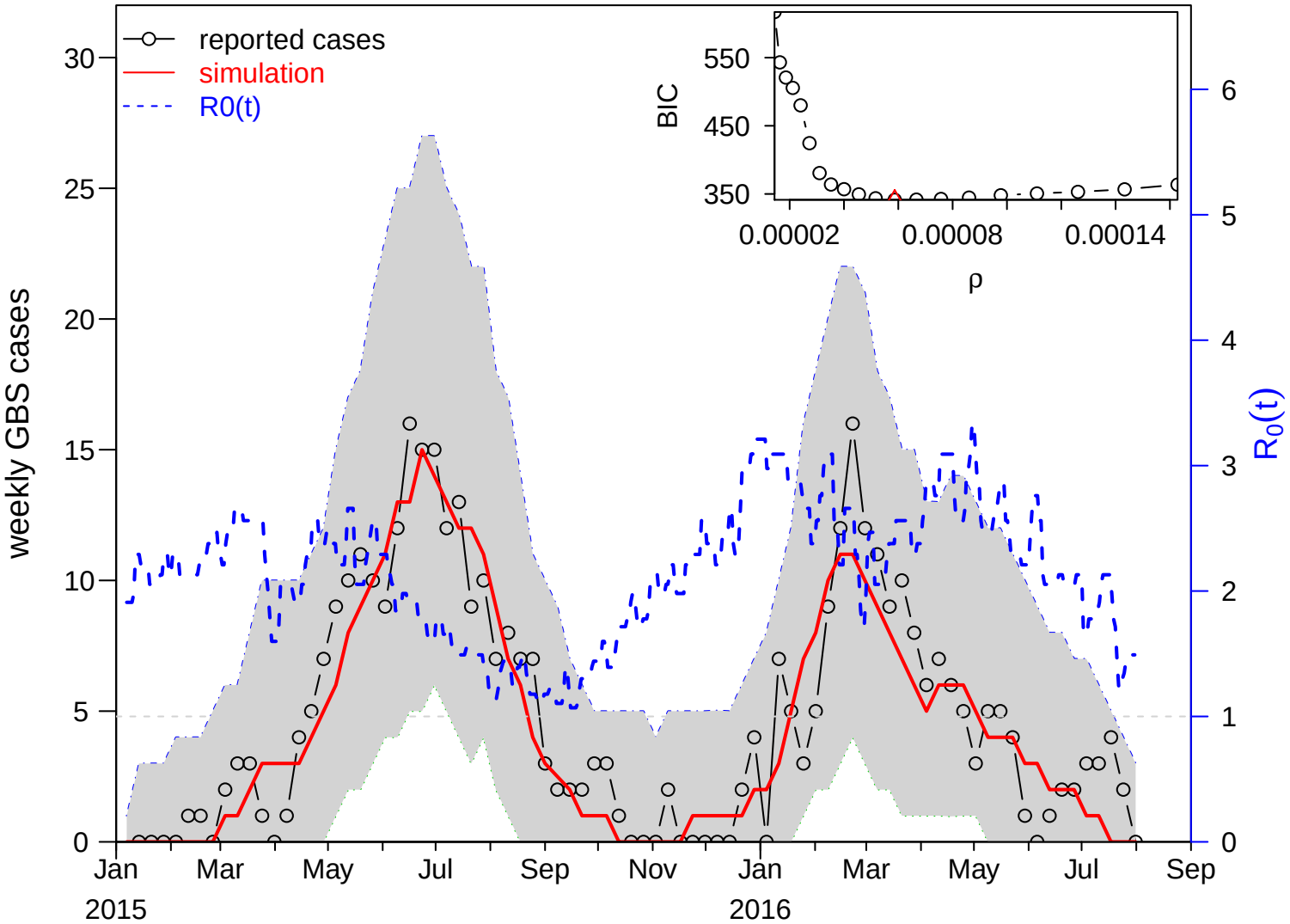


1918 Influenza in England (mechanistic model)

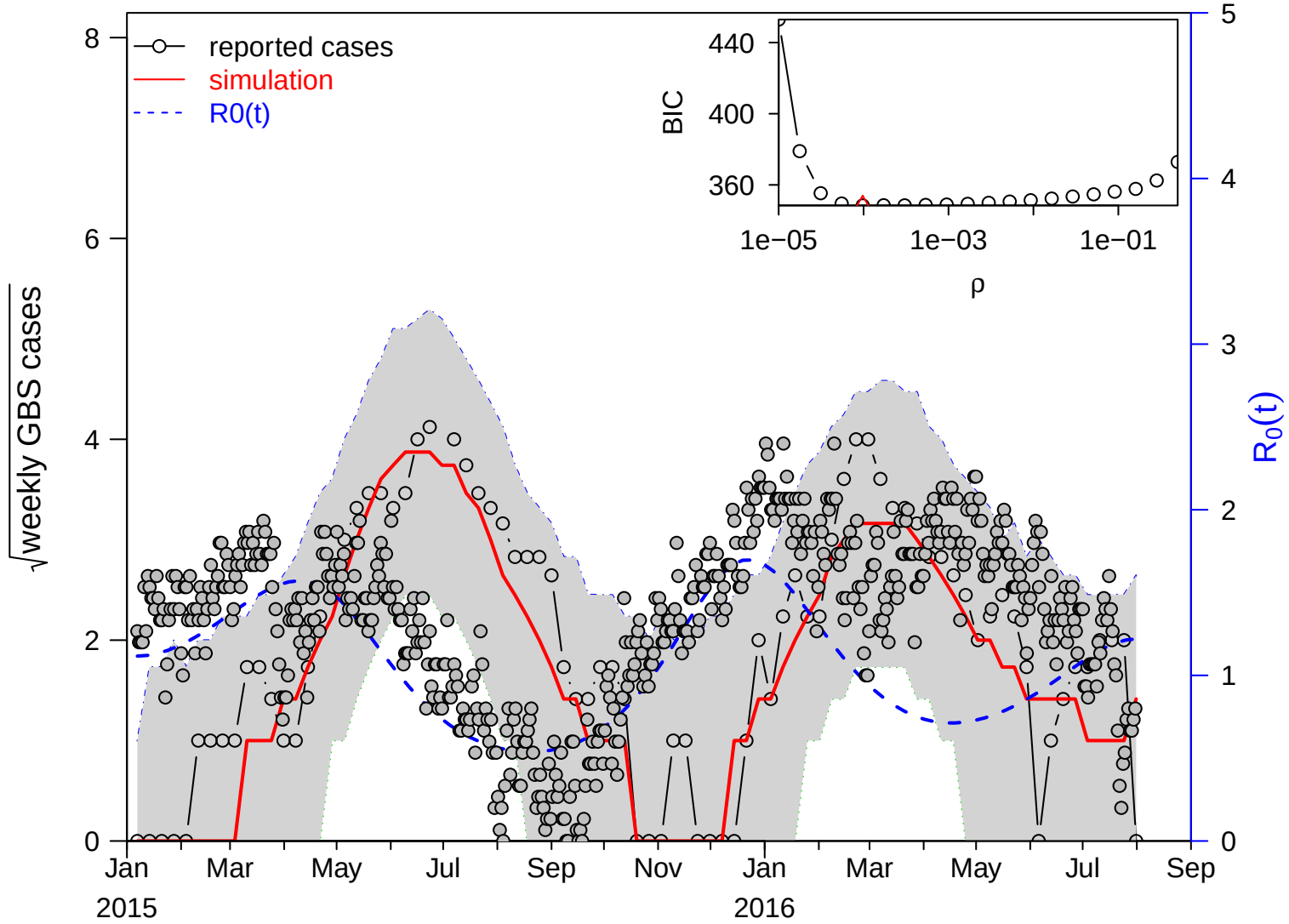


Outline

Zika fever in Brazil (temperature-driven)

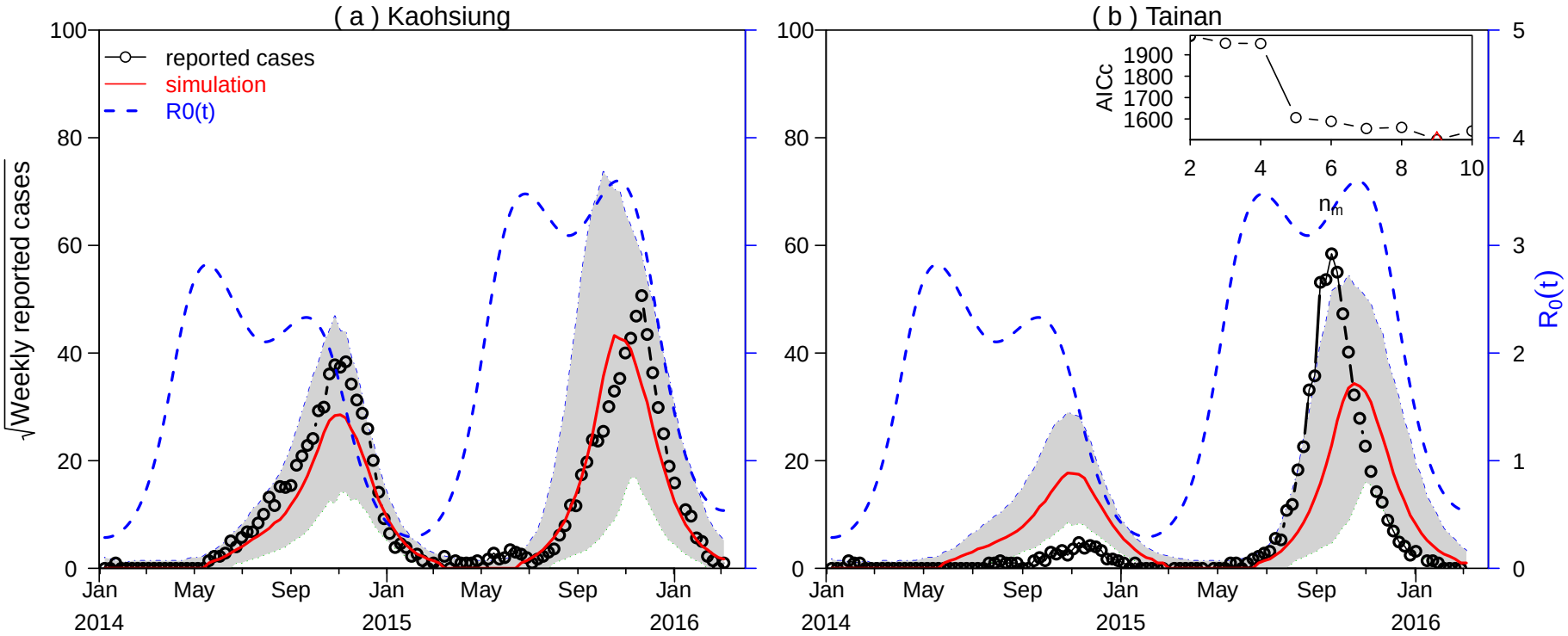


Zika fever in Brazil (non-mechanistic)

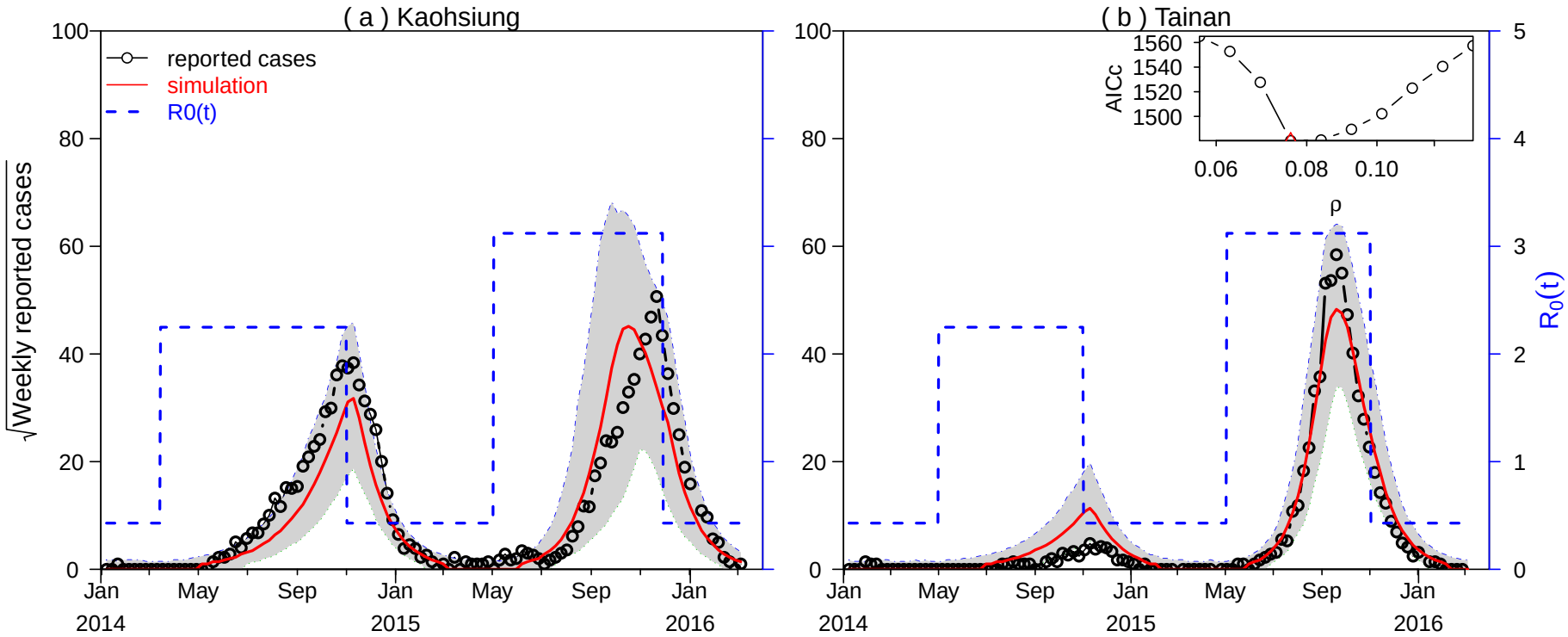


Outline

Dengue fever in Kaohsiung and Tainan



Dengue Fever in Kaohsiung and Tainan



Outline

Conclusions

- Plug-and-play likelihood-based inference framework is a powerful tool for disease modeling.
- While mechanistic models require additional assumptions and knowledge, non-mechanistic models do not.
- The comparison between fitting mechanistic models and fitting non-mechanistic models to data helps gain insight on possible mechanisms underlying disease outbreaks.

THANK YOU !

Reference


Ho S, He D, Eftimie R (2019) Mathematical models of transmission dynamics and vaccine strategies in Hong Kong during the 2017 - 2018 winter influenza season. Journal of Theoretical Biology. 476:74-94.


Lou Y, Liu K, He D, Gao D, Ruan S (2019) Modelling diapause in mosquito population growth. Journal of Mathematical Biology. 78(7): 2259-2288.


Musa S, Zhao S, Chan H, Jin Z, He D (2019) A mathematical model to study the 2014-2015 large-scale dengue epidemics in Kaohsiung and Tainan cities in Taiwan, China. Mathematical Biosciences and Engineering. 16(5): 3841-3863.


Lou Y, Liu K, He D, Gao D, Ruan S (2019) Modelling diapause in mosquito population growth. Journal of Mathematical Biology. 2019 Mar 1. doi: 10.1007/s00285-019-01343-6


Zhao S, Bauch C, He D (2018) Strategic Decision-Making about Travel during Disease Outbreaks: a Game Theoretical Approach. Journal of the Royal Society Interface. 15(146):20180515.


He D, Wang X, Gao D, Wang J (2018) Modeling the 2016-2017 Yemen Cholera Outbreak with the Impact of Limited Medical Resources. Journal of Theoretical Biology. 451:80-85.


Zhao S, Lou Y, Chiu A, and He D (2018) Modeling the Skip-and-Resurgence of Japanese Encephalitis Epidemics in Hong Kong. Journal of Theoretical Biology. 454:1-10.


Zhao S, Stone L, Gao D, and He D (2018) Modelling the Large-scale Yellow Fever Outbreak in Luanda, Angola, and the Impact of Vaccination. PLOS Neglected Tropical Diseases. 12(1): e0006158. Awards for Outstanding Research Articles in Biosurveillance


Lin Q, Chiu A, Zhao S and He D (2018) Modelling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia. Statistical Methods in Medical Research. 27(7):1968-1978.


Yu D, Lin Q, Chiu A, and He D (2017) Effects of Reactive Social Distancing on the 1918 Influenza Pandemic. PLOS One 0180545.


Tang X, Zhao S, Chiu A, Ma H, Xie X, Mei S, Kong D, Qin Y, Chen Z, Wang X, He D (2017) Modelling the transmission and control strategies of varicella among school children in Shenzhen, China. PLOS One. e0177514


He D, Gao D, Lou Y, Zhao S, and Ruan S (2017) A comparison study of Zika virus outbreaks in French Polynesia, Colombia and the State of Bahia in Brazil. Scientific Reports 7:273.


He D, Chiu A, Lin Q, Yu D (2017) Spatio-temporal patterns of proportions of influenza B cases. Scientific Reports 7: 40085.


Wang X, Yang L, He D, Chiu A, Chan K, Chan K, Guo Q, Zhou M, Wong C, Hu W (2017) Different Responses of Influenza Epidemic to Weather Factors among Shanghai, Hong Kong and British Columbia. International Journal of Biometeorology 61(6):1043-1053.


He D and Earn D (2016) The cohort effect in childhood disease dynamics. Journal of The Royal Society Interface 13, 20160156.


Gao D, Lou Y, He D, Porco T, Kuang Y, Chowell G, and Ruan S (2016) Prevention and control of Zika as a mosquito-borne and sexually transmitted disease: A mathematical modeling analysis. Scientific Reports 6, 28070 doi:10.1038/srep28070


Lin Q, Lin Z, Chiu A and He D (2015) Seasonality of Influenza A(H7N9) Virus in China - Fitting Simple Epidemic Models to Human Cases. PLOS One 0151333.


He D, Lui R, Wang L, Tse C, Yang L, and Stone L (2015) Global Spatio-temporal Patterns of Influenza in the Post-pandemic Era. Scientific Reports. 5:11013. doi: 10.1038/srep11013.


He D, Dushoff J, Eftimie R, and Earn D (2013) Patterns of Spread of Influenza A across Canada. Proc Roy Soc B, 280(1770), 20131174.


He D, Dushoff J, Day T, Ma J, and Earn D (2013) Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. Proc Roy Soc B 280(1766), 20131345.


Earn D, He D, Loeb M, Fonseca K, Lee B, and Dushoff J (2012) Effects of school closure on incidence of pandemic influenza in Alberta, Canada. Annals of Internal Medicine 156: 173-181.


He D, Dushoff J, Day T, Ma J, and Earn D (2011) Mechanistic modelling of the three waves of the 1918 influenza pandemic. Theoretical Ecology 4: 283-288


He D, Ionides E, and King A (2010) Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. Journal of The Royal Society Interface 7: 271-283. (This paper was among the top ten most cited in the journal in 2010)


Breto C, He D, Ionides E, and King A (2009) Time series analysis via mechanistic models. Annals of Applied Statistics 3: 319-348.