Discussion of "An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China"

Tianjian Zhou[1] and Yuan Ji[*1]

[1]*Department of Public Health Sciences, The University of Chicago*

We congratulate Wang et al. for their nice work on the COVID-19 epidemic in China. As of April 16, 2020, COVID-19 has become a pandemic and is affecting over 200 countries and regions in all continents except Antarctica. Modeling of COVID-19 data is of great importance, because it can provide insight into the dynamics of the spread of SARS-COV-2, the virus that causes the COVID disease, and the effects of mitigation policies. Such insight is helpful for health workers and policy makers to evaluate potential interventions and make forecast about the future trend of the virus spread. This is exactly the aim of Wang et al. In addition, we appreciate the authors' effort in providing an R package `eSIR`, which facilitates the data analysis using the proposed model. In what follows, we comment on the modeling approach taken by Wang et al. and suggest future directions for this area of research.

1 The Proposed State-space SIR Model

Generally speaking, epidemic models can be categorized as *deterministic* models and *stochastic* models. Most epidemic models consider a partition of the population into different *compartments*. These compartments correspond to individuals at different stages of a disease epidemic, such as susceptible individuals (who do not have the disease but can be infected), infectious individuals (who have the disease and can infect others), and removed individuals (who had the disease and cannot be infected again or infect others). Each individual precisely belongs to one of these compartments, i.e., the compartments are mutually exclusive and exhaustive. The spread of the disease is described by the flow of people across different compartments.

Deterministic models characterize disease transmission through a set of differential equations. For example, the classical susceptible-infectious-removed (SIR) model (Kermack and McKendrick, 1927) is given by the following equations:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \text{and } \frac{d\theta_t^R}{dt} = \gamma\theta_t^I. \tag{1}$$

Here, following the notations in Wang et al., $\boldsymbol{\theta}_t = (\theta_t^S, \theta_t^I, \theta_t^R)^\top$ denotes the prevalences of susceptible, infectious and removed individuals in the entire population at time t, and β and γ are the transmission and removal rates, respectively. Given the initial prevalences of the three compartments and the parameter values, the trajectory of $\boldsymbol{\theta}_t$ over time is fully determined by Equations (1).

In practice, the spread of disease is rarely a deterministic process. Furthermore, it is likely that the prevalences of some compartments are observed with error or may not be observed at all. As a result, stochastic/statistical modeling is needed to take these considerations into account. The state-space SIR model developed by Osthus et al. (2017) incorporates measurement errors by modeling the observed prevalences as random variables centered at values indicated by a

---

*Corresponding author. Email: yji@health.bsd.uchicago.edu.

deterministic SIR model. Specifically,

$$Y_t^c \mid \boldsymbol{\theta}_t \sim \text{Beta}(\lambda^c \theta_t^c, \lambda^c(1 - \theta_t^c)), \quad \text{and} \quad \boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1} \sim \text{Dirichlet}(\kappa f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)). \tag{2}$$

Here, Y_t^c is the observed proportion of compartment $c \in \{S, I, R\}$ at time t, and $f(\cdot)$ is the deterministic SIR curve indicated by Equations (1). The deviation of Y_t^c from the deterministic SIR trajectory is captured by additional variance parameters λ^c and κ. Model (2) was originally developed in Osthus et al. (2017) for modeling seasonal influenza. The idea of model (2) is quite general and has also been used in other applications such as Osthus et al. (2019). Wang et al. extended this modeling framework, specifically for the analysis of COVID-19 data, by considering (i) a time-varying transmission rate, $\beta = \beta(t)$, or (ii) a quarantine compartment θ_t^Q in the model. In addition, the Runge-Kutta approximation of $f(\cdot)$, first proposed in Osthus et al. (2017) and then adopted by Wang et al., facilitates posterior computation.

In model (2), it can be seen that $\boldsymbol{\theta}_t$ is a (latent) discrete-time Markov process. An alternative way of specifying the process $\{\boldsymbol{\theta}_t, t \geq 0\}$ is through a stochastic SIR model directly. See, for example, Andersson and Britton (2000). Specifically, one may assume that the times each infectious individual has contacts with a given susceptible individual follows a Poisson process of rate β. Suppose any of these contacts can result in the susceptible individual being infectious immediately. Furthermore, suppose the time between the infection and removal of an infectious individual follows an exponential distribution of rate γ. Assume all those Poisson processes and exponential distributions are independent of each other. Then, $\boldsymbol{\theta}_t$ is a continuous-time Markov process. This direction may be considered in the future.

2 Posterior Inference for the State-space SIR Model

Wang et al. used a Markov chain Monte Carlo algorithm to obtain draws from the posterior distribution of the parameters. Specifically, the R package `rjags` (Plummer et al., 2019) was used for posterior simulation. We note that since model (2) is essentially a dynamic model, online and sequential algorithms such as sequential Monte Carlo (Doucet et al., 2001) may be considered in the future to improve the efficiency of posterior sampling. In that way, when data at more time points become available, one can update the posterior in an efficient way rather than re-fitting the model to the complete data.

It is also worth noting that parameter estimations of model (2) may be sensitive to the choice of hyperparameters. Wang et al. specified the hyperparameters in a sensible way using SARS data.

3 Concluding Remarks

The work by Wang et al. illustrates the potential and power of statistical modeling for infectious diseases like COVID-19. The work was focused on the COVID-19 outbreak in China, but the proposed model is also applicable to the data analysis for other countries like Italy or the United States. The COVID-19 pandemic is still affecting many countries and may not end in the near future. We hope to see more data-driven and model-based inference for the dynamics of the spread of COVID-19 and hope such inference could be helpful for policy makers and health workers.

References

Andersson H, Britton T (2000). *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer Science & Business Media, New York.

Doucet A, De Freitas N, Gordon N (2001). An introduction to sequential Monte Carlo methods. In: *Sequential Monte Carlo Methods in Practice* (A Doucet, N De Freitas, N Gordon, eds.), 3–14. Springer Science & Business Media, New York.

Kermack WO, McKendrick AG (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, 115(772): 700–721.

Osthus D, Gattiker J, Priedhorsky R, Del Valle SY (2019). Dynamic Bayesian influenza forecasting in the United States with hierarchical discrepancy (with discussion). *Bayesian Analysis*, 14(1): 261–312.

Osthus D, Hickmann KS, Caragea PC, Higdon D, Del Valle SY (2017). Forecasting seasonal influenza with a state-space SIR model. *The Annals of Applied Statistics*, 11(1): 202–224.

Plummer M, Stukalov A, Denwood M (2019). rjags: Bayesian graphical models using MCMC. https://cran.r-project.org/web/packages/rjags/.