Invited Discussion

Vladimir N. Minin[*], Jonathan Fintzi[†], Luis J. Martinez Lomeli[‡], and Jon Wakefield[†,§]

The authors present an elegant method for accurate prediction of influenza-like-illness (ILI) incidence during an ongoing flu season. Their method combines ordinary differential equation-based (ODE-based) mechanistic modeling of ILI spread with flexible modeling of discrepancies between the ODE trajectories and observed incidence. The key idea is that these discrepancies behave similarly across flu seasons. Capturing these similarities in a Bayesian hierarchical model, the authors arrive at a predictive semiparametric model of ILI spread. The authors conjecture that there is room for improving their approach and discuss some enhancements to the nonparametric component of their model. Below we argue that more careful handling of the parametric model component may also be a fruitful strategy to pursue in parallel with nonparametric model enhancements.

Flexible modeling and forecast sharpness

The authors motivate their discrepancy model component by correctly pointing out that certain consistently repeated features of ILI incidence time series cannot be predicted using deterministic mechanistic epidemic models. The authors' results show that the new Bayesian hierarchical model can indeed capture these features. For example, Figure 7 in the Osthus et al. manuscript shows that a consistent, but mysterious drop in ILI incidence from week 13 to week 14 can be seen in the authors' short term forecasts. However, the same figure shows that using the first 4 and 8 weeks of ILI data produces weeks 10-25 predictive intervals that are so large that they cover almost the entire plausible range of weighted ILI (wILI) counts. This suggests that the authors' model may be a little too flexible. There are multiple ways to tighten the authors' model, but from our perspective, the most intriguing avenue to pursue is to try to improve the parametric model component. Specifically, we first concentrate modeling efforts on improving the mean model, to reduce bias. Second, we finesse the wILI variance model, in particular paying attention to how the variance depends on the mean, so that we obtain an appropriate measure of uncertainty.

SIR-only model

Incidence ODEs with unknown initial conditions To establish a baseline, we wanted to see how SIR-only predictions compare to the authors' much more advanced modeling. Following the authors, we model the transmission dynamics of wILI in the population

[*] Department of Statistics, University of California, Irvine, CA, vminin@uci.edu
[†] Department of Biostatistics, University of Washington, Seattle, WA
[‡] Center for Complex Biological Systems, University of California, Irvine, CA
[§] Department of Statistics, University of Washington, Seattle, WA

using a Susceptible-Infected-Recovered (SIR) model, represented as a system of ODEs. Let $\mathbf{X}^{(j)}(t) = (S^{(j)}(t), I^{(j)}(t), R^{(j)}(t))$, $S^{(j)}(t)+I^{(j)}(t)+R^{(j)}(t) = K$, denote the vector of compartment counts at time t in season $j \in \{1998, \ldots, 2014\}$, where K is the population size that we set to 3×10^8 to approximate the size of U.S. population. We also let $\mathbf{X}_0^{(j)} = (S_0^{(j)}, I_0^{(j)}, R_0^{(j)})$ be the initial compartment counts.

The standard ODE representation of the SIR model expresses the time-evolution of the compartment counts as the solution to the following system of ODEs:

$$\begin{aligned}
&\frac{\mathrm{d}S^{(j)}(t)}{\mathrm{d}t} = -\beta_j S^{(j)}(t) I^{(j)}(t), \;\; \frac{\mathrm{d}I^{(j)}(t)}{\mathrm{d}t} = \beta_j S^{(j)}(t) I^{(j)}(t) - \gamma_j I^{(j)}(t), \qquad (1)\\
&\frac{\mathrm{d}R^{(j)}(t)}{\mathrm{d}t} = \gamma_j I^{(j)}(t), \text{ such that, } \mathbf{X}^{(j)}(0) = \mathbf{X}_0^{(j)},
\end{aligned}$$

where β_j is the per–contact infection rate in season j and γ_j is the recovery rate. This is the same model that the authors use as their parametric component.

We modify the authors' SIR model in two ways. First, we are skeptical of the authors' claim that the initial number of susceptible individuals in each season is not identifiable. This claim may be true if only one season/outbreak is observed, but availability of multiple season onsets can make the initial number of susceptibles identifiable. To explore this issue, we introduce an additional parameter, C_j, for the number of susceptibles who are effectively removed at the start of season j, e.g., due to pre–existing immunity or geographic isolation. Second, to make the SIR model more appropriate for the incidence data, we follow Bretó and Ionides (2011) and Ho et al. (2018) and reparameterize the SIR ODEs in terms of cumulative incidence. Let $\mathbf{N}^{(j)}(t) = (N_{SI}^{(j)}(t), N_{IR}^{(j)}(t))$ denote the cumulative numbers of infections and recoveries and $\mathbf{N}^{(j)}(0)$ be the initial numbers of these events. The SIR ODEs for cumulative incidence and recoveries are given by

$$\begin{aligned}
\frac{\mathrm{d}N_{SI}^{(j)}(t)}{\mathrm{d}t} &= \beta_j \left(S_0^{(j)} - C_j - N_{SI}^{(j)}(t)\right)\left(I_0^{(j)} + N_{SI}^{(j)}(t) - N_{IR}^{(j)}(t)\right), \qquad (2)\\
\frac{\mathrm{d}N_{IR}^{(j)}(t)}{\mathrm{d}t} &= \gamma_j \left(I_0^{(j)} + N_{SI}^{(j)}(t) - N_{IR}^{(j)}(t)\right), \mathbf{N}^{(j)}(0) = (0,0).
\end{aligned}$$

Notice that we need the initial compartment counts $\mathbf{X}_0^{(j)}$ in the above system. Technically, we do not need to have both C_j and R_0^j in our model, because they represent the same number of initially removed individuals. We set $R_0^j = 0$ and estimate C_j due to constraints of our pre-baked implementation of the SIR model.

We fit two versions of our modified model to 15 seasons corresponding to years 1998–2007 and 2010–2014. In the first model A we assume that $C_j = C$, for j, with C being an unknown parameter that we estimate together with season-specific infection and recovery rates. We use this model primarily to test whether C is identifiable. The second model B is more realistic and assumes that each season j can have its own number of initially removed individuals, C_j. The model is hierarchical in that it assumes that *a priori* C_j's are drawn independently from the same distribution. More specifically, $\mathrm{logit}(C_j/K) \sim \mathcal{N}(\mu_C, \sigma_C^2)$, with unknown parameters μ_C and σ_C^2 that we estimate.

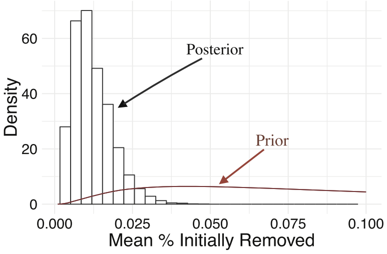

Figure 1: Prior density and posterior histogram of the proportion of initially removed individuals C/K. The prior and posterior are for the simple model in which all seasons start with the same number of initially removed individuals.

Data model Let $P_{SI}^{(j)}(t) = N_{SI}^{(j)}(t)/K$ be the attack rate (% of the population infected) up to time t in season j. Let $\Delta P_{SI}^{(j)}(t_\ell) = P_{SI}^{(j)}(t_\ell) - P_{SI}^{(j)}(t_{\ell-1})$ denote the attack rate in week ℓ. We model the observed wILI in week ℓ of season j, denoted $Y_\ell^{(j)}$, as

$$\operatorname{logit}\left(Y_\ell^{(j)}\right) \sim \mathcal{N}\left(\operatorname{logit}\left(\Delta P_{SI}^{(j)}(t_\ell)\right), \frac{\omega_0 + \omega_1 \Delta P_{SI}^{(j)}(t_\ell)}{\Delta P_{SI}^{(j)}(t_\ell)\left(1 - \Delta P_{SI}^{(j)}(t_\ell)\right)^2}\right), \tag{3}$$

where ω_0 and ω_1 control the variance of the emission distribution. This measurement model derives from an application of the delta method to a normal approximation of an overdispersed binomial distribution for detected wILI cases under the assumption that the rate of patient visits is not changing across time. The main motivation for this fairly complicated data model is our desire to model the dependence of wILI count variance on the latent/unobserved population incidence.

Priors and posterior inference We assign informative, scientifically meaningful priors, detailed in Table 1, for the parameters of models A and B. Note that we assign Dirichlet-Multinomial prior to the initial state $\mathbf{X}_0$ in such a way that there are no removed individuals at time 0, because we have a separate parameter to the number of initially removed individuals, C_j. For model A, where all parameters but the number of removed individuals C, are decoupled across all the seasons, we used our custom Markov chain Monte Carlo (MCMC) algorithm to approximate the posterior distribution of $(\beta_j, \gamma_j, I_0^j, S_0^j, \omega_0, \omega_1)$ for each season j and C that is common to all seasons. We show the prior and posterior distributions of the number of removed individuals C in Figure 1. The apparent differences between the prior and posterior distributions suggests that parameter C is identifiable. Moreover, our proportion of initially removed individuals C/K is much lower than 0.1 — the number used by Osthus et al. In Model B, we used MCMC to target the posterior distribution of $(\beta_j, \gamma_j, I_0^j, S_0^j, C_j, \omega_0, \omega_1)$ for each season j and (μ_C, σ_C). We found that the season-specific C_j's and their overall prior mean μ_C and standard deviation σ_C were also identifiable. We omit most of posterior summaries for the sake of brevity.

Model	**Parameter**	**Interpretation**	**Prior**	**Prior Median (90% Interval)**
A	$R0^{(j)} = \beta_j(K - C)/\mu_j - 1$	Basic reproduction #-1	LogNormal(log(0.4), 1.25)	$R0^{(j)} = 1.4$ (1.05, 4.10)
B	$R0^{(j)} = \beta_j(K - C_j)/\mu_j - 1$	Basic reproduction #-1	LogNormal(log(0.4), 1.25)	$R0^{(j)} = 1.4$ (1.05, 4.10)
A,B	$7/\mu_j - 1$	Mean infectious period (days-1)	LogNormal(log(7), 0.843)	$7/\mu_j = 7$ (1.75, 28)
A	C/K	% initially removed	LogitNormal(logit(0.1), 1)	$C/K = 0.1$ (0.02, 0.37)
B	μ_C	Mean logit % initially removed	LogitNormal(logit(0.1), 0.63)	expit(μ_C) = 0.1 (0.04, 0.24)
B	σ_C	Std.dev. logit % initially removed	Exponential(4)	$\sigma_C = 0.17$ (0.013, 0.75)
B	C_j/K	% initially removed	LogitNormal(μ_c, σ_C^2)	—
A,B	ω_0	Variance parameter	Exponential(3×10^8)	$\omega_0 = 2.3 \times 10^{-9}$ (1.7×10^{-10}, 1.0×10^{-8})
A,B	ω_1	Variance parameter	Exponential(5)	$\omega_1 = 0.14$ (0.01, 0.60)
A,B	$\mathbf{X}_0$	Initial compartment counts	Dirichlet–Multinom.(150,2,0)	

Table 1: Parameters and priors used in fitting SIR models to wILI data. Under Model A, the initial depletion of susceptibles is common to all seasons, whereas Model B hierarchically allows for season–specific initial depletion of susceptibles.

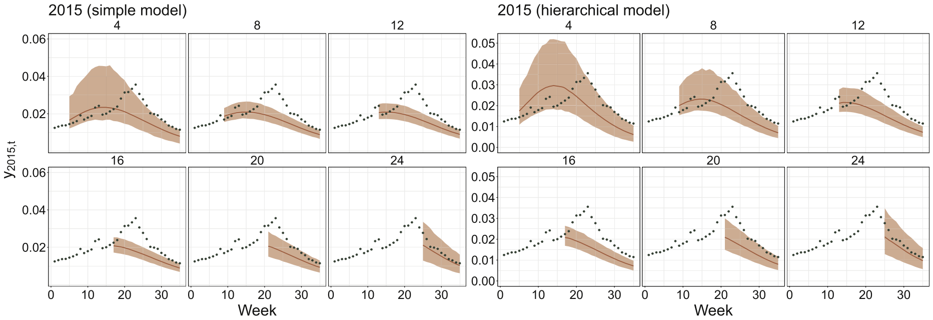

Figure 2: Forecasts for 2015 season under models A and B. The left six plots show forecasts produced by model A, where all seasons share the same number of initially removed individuals. The right six plots show forecasts produced by model B, where this number of initially removed individuals is season-specific. Each plot has the first z weeks/points used as training data, with the rest of the data being withheld during model fitting. This number z is shown above each plot (e.g., $z = 4$ in the top left plot). The solid red lines show the medians of the predictive distributions on which the forecasts are based. The shaded areas designate 95% predictive intervals.

SIR-only predictions

Now we use our SIR models A and B to make predictions about wILI incidence in season 2015. During this forecasting exercise, we use the estimated posterior distributions of SIR model parameters for seasons 1998–2007 and 2010–2014 in the following way. We pool MCMC samples of season-specific parameters and fit a multivariate Gaussian mixture model to these samples. For model A, separately from the mixture model fitting, we approximate the posterior distribution of initially removed individuals C with a univariate log-normal distribution. We use these approximations to the posterior model parameters as priors in our analysis of partial data from the 2015 season. As in the authors' paper, we fit our SIR model A to the first z weeks of data and use this model to predict the rest of the season for $z = 4, 8, 12, 16, 20, 24$. Prediction results are shown in Figure 2. Both sets of priors result in reasonable short term forecasts in weeks 4, 8, 12, and 24, but the timing of the epidemic peak is not predicted well. We see that a mixture model-based prior distribution of the initial number of removed individuals, obtained from the posterior samples under model B, produces better forecasts than predictions based on a prior distribution obtained from the posterior of model A. However, similarly to the Osthus et al. hierarchical model, this improvement comes at the expense of wider predictive intervals. Still, our experiments with the initial states of flu seasons demonstrate that careful modeling of initially removed individuals may be a fruitful forecasting strategy, at least in the context of time homogeneous infectious disease dynamics.

Speculative remarks

Although our SIR-only predictions are not competitive with the state-of-the-art ILI forecasting methods, they establish a parametric modeling starting point, which is different from the starting point of Osthus et al. Combining parametric modeling similar to ours with the authors' hierarchical discrepancy model may improve ILI forecasting even further. More specifically, it would be interesting to see if including the initial number of removed individuals as a free parameter and/or a data model with a mean/variance relationship into Osthus et al. model would lead to better forecasts.

Another way to improve SIR-only predictions is to use stochastic SIR modeling and to move to a nonparametric modeling of the infection rate, as was recently proposed by Xu et al. (2016) in a wider context of stochastic epidemic modeling. For example, we can assume that for season each j the time-varying infection rate has the form $\beta_j(t) = \alpha_j \times \beta(t)$, where α_j's are season-specific multipliers and $\beta(t)$ captures commonalties in infection rate changes across seasons. *A priori* modeling of $\beta(t)$ as a Gaussian process or another suitable functional prior would result in nonparametric estimation of $\beta(t)$. In summary, we are excited about the successes of Osthus et al. forecasting method based on semi-parametric modeling of infectious disease dynamics and looking forward to future modeling and forecasting improvements in this area.

References

Bretó, C. and Ionides, E. (2011). "Compound Markov counting processes and their applications to modeling infinitesimally over–dispersed systems." *Stochastic Processes and their Applications*, 121: 2571–2591. MR2832414. doi: `https://doi.org/10.1016/j.spa.2011.07.005`. 302

Ho, L., Crawford, F., and Suchard, M. (2018). "Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease." *The Annals of Applied Statistics*, 12: 1993–2021. MR3852706. doi: `https://doi.org/10.1214/18-AOAS1141`. 302

Xu, X., Kypraios, T., and O'Neill, P. (2016). "Bayesian non-parametric inference for stochastic epidemic models using Gaussian processes." *Biostatistics*, 17(4): 619–633. MR3604269. doi: `https://doi.org/10.1093/biostatistics/kxw011`. 306