Real-Time Forecasting of Seasonal Influenza in South Korea with Compartment model and Assimilation Filtering

Minhye Kim * **Kyeongah Nah** ** **Masud M A** *** **Sangil Kim** **** **Yongkuk Kim** *

* *Department of Mathematics, Kyungpook National University, Daegu, Korea*
** *Busan Center for Medical Mathematics, National Institute for Mathematical Sciences, Busan, Korea (e-mail: knah@nims.re.kr)*
*** *Natural Product Informatics Research Center, Korea Institute of Science and Technology, Gangneung, Korea*
**** *Department of Mathematics, Pusan National University, Busan, Korea*

Abstract: Seasonal influenza is an acute respiratory infection caused by several types of influenza viruses worldwide. Its outbreak exhibits a seasonal cycle in temperate climates. For public health decision-making and medical resource management during the time course of seasonal epidemics, a reliable real-time forecasting system is necessary. In this study, we introduce a novel approach combining two different data assimilation techniques to produce a real-time prediction of seasonal influenza governed by the standard SIR model. When applying our developed approach to Influenza-Like-Illness(ILI) data collected in Korea for 2016–2021, it successfully near-casted the upcoming week's flu incidence.

Keywords: Variational data assimilation, Particle filtering/Monte Carlo methods, Modeling and identification, Epidemic modeling.

1. INTRODUCTION

Influenza-like-illness is an acute respiratory infection appearing with cough and fever. As it shows up in seasonal cycles in South Korea, supporting the infected individuals with proper treatment and care using the existing medical resources becomes a challenge during the outbreak.

Data fitting techniques along with cutting edge machine learning algorithms are commonly used to forecast epidemic trend of an outbreak. During Covid-19 pandemic, several machine learning techniques approaches are raised and provided fairly accurate prediction of the epidemic trends [Wang et al. (2020); Tuli et al. (2020)]. In general, the machine learning models are not interpretable from public health perspective and hence comes with limitation in applicability. Instead, SIR-type of compartment models which incorporates the mechanisms of disease transmission and controls are often used in analyses which requires specific interpretation, such as in scenario analysis [Osthus et al. (2017); Law et al. (2021); Bjørnstad et al. (2002); Law et al. (2020)]. Although SIR-type of models resolve the issue of interpretablity, it is often studies with constant values of the parameters over the whole period of interest which however changes with time due to many associated factors like environmental conditions, public health measures, etc. Markov Chain Monte Carlo derived techniques though produce time dependent estimates, its accuracy is low in the beginning. To address this issue, we use variational data assimilation technique to predict the initial states of SIR model and then use Bayesian filtering to forecast the flu incidence of the upcoming weeks.

2. METHODS

We assume the following simple SIR model with constant population $N = S(t) + I(t) + R(t)$.

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)\frac{I(t)}{N} \\ \frac{dI(t)}{dt} &= \beta S(t)\frac{I(t)}{N} - \sigma I \\ \frac{dR(t)}{dt} &= \sigma I \end{aligned} \tag{1}$$

where, β is the transmission rate and σ is the recovery rate. Firstly, we estimate β and initial states using ICC curves and data assimilation techniques. As introduced in Lega (2021), for a total susceptible population of size N and each initial condition $k = \frac{S(0)}{N} = 1 - \frac{C(0)}{N}$, the ICC curve of the model (1) is given by,

$$G_{k,N} = \beta\left(C + \frac{N}{R_0}\ln\left(1 - \frac{C}{N}\right) - \frac{N}{R_0}\ln(k)\right)\left(1 - \frac{C}{N}\right) \tag{2}$$

where, $C = I + R \in [0, C_\infty]$ and C_∞ the final number of cases is the positive solution of the transcendental equation

$$C_\infty + \frac{N}{R_0}\ln\left(1 - \frac{C_\infty}{N}\right) - \frac{N}{R_0}\ln(k) = 0 \tag{3}$$

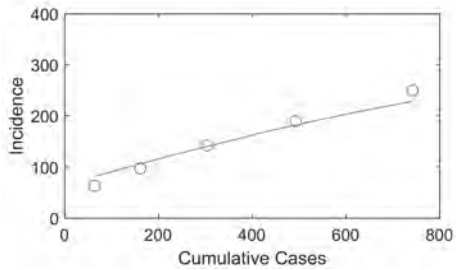

Fig. 1. Blue circles represent the pairs of cumulative cases and incidences of ILI data. Red curve is SIR model-derived ICC curve with parameters that best fits the cumulative cases – incidence data.

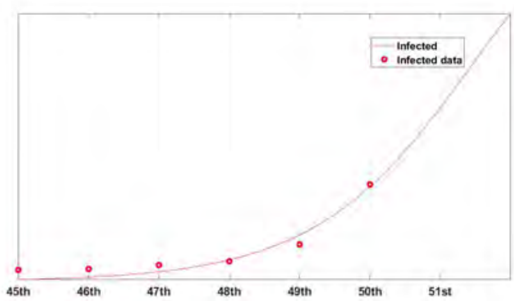

Fig. 2. Predicted numbers of weekly ILI cases using VDA method based on SIR model.

where R_0 is the basic reproduction number, which is one of the important quantities in dynamics, and is obtained as $R_0 = \beta/\sigma$. It is the expected number of secondary cases generated by one case in a population where no other individuals are infected or immunized.

We estimate the values of $C(0)$ and σ using cumulative incidence data of the first week and empirically observed duration of infectiousness, respectively. Then, the unknown value of the transmission rate β is determined to be the one which best-fit ICC to the observed incidence data, see Fig. 1.

Next, using the Variational data assimilation(VDA) introduced by Rhodes and Hollingsworth (2009), we find the initial states of SIR model which yield the best-fit of the model to the observed data. Fig. 2 shows that the number of ILI patients at 51 weeks can be predicted through initial state estimates. In the numerical optimization process, we used the adjoint method to compute gradients of the predefined cost function. In order to forecast the flu incidence in real-time, we adapt the technique of Bayesian filtering. Starting with the estimated initial states, we use Particle filtering in each time step to forecast the states in the next time step. Upon receiving the observed data in the next time step the forecast is updated to produce the estimate, which is further used to forecast in the next time step. And the cycle goes on until the end.

3. RESULTS

We implement the above mentioned approach to weekly incidence ILI data of South Korea over the period 2016-2021. The Fig. 3 shows good agreement of forecast with the trend observed in the data for two consecutive peaks

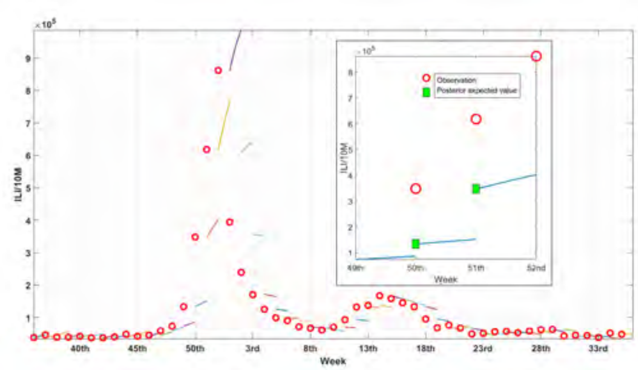

Fig. 3. Weekly forecasting of ILI cases. The red circles represent the weekly data and the solid lines shows weekly forecast.

during 2016-2017. A limitation of the suggested technique is that there is no established method in determining the time of season onset and it has to be changed each year for the accurate prediction of the epidemic curve. As a future work, we aim to develop a method which determines the time of season onset using Markov Switching.

REFERENCES

Bjørnstad, O.N., Finkenstädt, B.F., and Grenfell, B.T. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological monographs*, 72(2), 169–184.

Law, K.B., M Peariasamy, K., Mohd Ibrahim, H., and Abdullah, N.H. (2021). Modelling infectious diseases with herd immunity in a randomly mixed population. *Scientific Reports*, 11(1), 1–12.

Law, K.B., Peariasamy, K.M., Gill, B.S., Singh, S., Sundram, B.M., Rajendran, K., Dass, S.C., Lee, Y.L., Goh, P.P., Ibrahim, H., et al. (2020). Tracking the early depleting transmission dynamics of covid-19 with a time-varying SIR model. *Scientific reports*, 10(1), 1–11.

Lega, J. (2021). Parameter estimation from ICC curves. *Journal of Biological Dynamics*, 15(1), 195–212.

Osthus, D., Hickmann, K.S., Caragea, P.C., Higdon, D., and Del Valle, S.Y. (2017). Forecasting seasonal influenza with a state-space SIR model. *The annals of applied statistics*, 11(1), 202.

Rhodes, C. and Hollingsworth, T. (2009). Variational data assimilation with epidemic models. *Journal of Theoretical Biology*, 258(4), 591–602.

Tuli, S., Tuli, S., Tuli, R., and Gill, S.S. (2020). Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing. *Internet of Things*, 11, 100222.

Wang, P., Zheng, X., Li, J., and Zhu, B. (2020). Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 139, 110058.