Infectious Disease Modelling 8 (2023) 471–483

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

A discrete-time susceptible-infectious-recovered-susceptible model for the analysis of influenza data

Georges Bucyibaruta [a, *], C.B. Dean [a], Mahmoud Torabi [b]

[a] Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave W, Waterloo, ON, N2L 3G1, Canada
[b] Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, R3E 0W3, Canada

ARTICLE INFO

Article history:
Received 2 June 2021
Received in revised form 29 April 2023
Accepted 29 April 2023
Available online 6 May 2023
Handling Editor: Dr. Jianhong Wu

Keywords:
Discrete-time epidemic model
Infectious diseases
Influx process
Non-linear stochastic dynamics
Seasonal influenza
SIRS model
Transmission parameter

ABSTRACT

We develop a discrete time compartmental model to describe the spread of seasonal influenza virus. As time and disease state variables are assumed to be discrete, this model is considered to be a discrete time, stochastic, Susceptible-Infectious-Recovered-Susceptible (DT-SIRS) model, where weekly counts of disease are assumed to follow a Poisson distribution. We allow the disease transmission rate to also vary over time, and the disease can only be reintroduced after extinction if there is a contact with infected individuals from other host populations. To capture the variability of influenza activities from one season to the next, we define the seasonality with a 4-week period effect that may change over years. We examine three different transmission rates and compare their performance to that of existing approaches. Even though there is limited information for susceptible and recovered individuals, we demonstrate that the simple models for transmission rates effectively capture the behaviour of the disease dynamics. We use a Bayesian approach for inference. The framework is applied in an analysis of the temporal spread of influenza in the province of Manitoba, Canada, 2012–2015.

1. Introduction

Seasonal influenza or flu is a respiratory infection caused by the influenza virus. Seasonal influenza has become a source of considerable human morbidity and mortality, which makes it a crucial example of a persisting and recurrent disease, especially in individuals who are older, pregnant, immunocompromised, or have chronic underlying disease (e.g. heart disease, diabetes, cancer, respiratory illnesses). Prior to our covid-transformed environment, an estimated 5%–10% of the Canadian population would become infected with influenza each year, with the highest rate occurring in children (PHAC, 2012). In addition, serious illness and death related to influenza occur more frequently in older adults (> 65 years) and persons with underlying medical conditions. As with most acute viral respiratory infections, seasonal influenza occurs annually in the fall and winter seasons with community outbreaks. The exact timing and duration (onset, peak, and end) of influenza activity vary and cannot be predicted precisely from one season to the next, but influenza activity often starts to increase in October, peaking between December and February and lasting as late as May (PHAC, 2012).

* Corresponding author.
E-mail addresses: gbucyibaruta@gmail.com (G. Bucyibaruta), cdean@uwaterloo.ca (C.B. Dean), Mahmoud.Torabi@umanitoba.ca (M. Torabi).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2023.04.008

The 2019 novel coronavirus disease (COVID-19) epidemic that originated in Wuhan, China, has affected more than 219 countries and territories around the world and resulted in 117 538 168 confirmed cases and 2 607 030 deaths up to 08 March 2021 (WI, 2021). During early days, the transmission characteristics appeared to be of similar magnitude to severe acute respiratory syndrome-related coronavirus (SARS-CoV) and pandemic influenza, indicating a risk of global spread. The course of the COVID-19 pandemic has changed the dynamics of the spread of influenza in marked and unexpected ways. A dramatic decrease in influenza cases has been noted in Canada most likely because of the public health interventions put in place to combat COVID-19 infection (PHAC, 2021). However, the development of general models to understand the dynamics of seasonal influenza is important in the long term.

Since size and timing of the outbreak are not deterministic but random events, with many unknown drivers, epidemic dynamics pose a great challenge to stochastic modelling (Finkenstadt et al., 2002). The popular mathematical framework for describing the epidemic dynamic in humans is known as continuous time compartmental models which are mathematically described by the so-called Susceptible-Infected-Recovered (SIR) models. These models assume that a population of interest is divided into distinct compartments according to the disease status: Susceptible (if previously unexposed to the infectious disease), Infected (if currently colonized by the infectious disease), and Recovered (if the infection is successfully cleared) and describes the evolution of the infectious diseases through changes in the number of individuals in each compartment over time (Keeling & Rohani, 2008).

Infected individuals recover with immunity but after some time they become susceptible again because of the rapidly evolving nature of the influenza virus (Yaari et al., 2013). Hence, to control a disease with a relatively long course of infection, or a disease with temporary immunity that has already become endemic, like influenza, the recruitment of susceptible individuals from loss of immunity cannot be neglected.

Previous efforts have proposed to model and understand the likely spatial spread of influenza. Lawson and Song (2010) developed a SIR model and applied it to publicly available influenza data at the county level considering only one flu season. Yaari et al. (2013) proposed and fitted a stochastic, discrete-time, age-of-infection Susceptible-Infectious-Recovered-Susceptible (SIRS) seasonal model to surveillance data in order to study seasonal influenza by incorporating antigenic drift into the model under the maximum-likelihood (frequentist) approach.

To address both incorporation of heterogeneous populations and modeling of the spread of an infectious disease, mathematical models have been considered at both individual and area levels (Deeth & Deardon, 2016; Lawson, 2018; Lawson & Song, 2010). These models provide a quantitative framework in which scientists can assess hypotheses on the potential underlying mechanisms that explain patterns in observed data at different temporal scales, and generate estimates of key parameters and forecasts.

Individual-level models provide an accurate description of the spread of infectious diseases in a population (Deeth & Deardon, 2016). In addition, they allow for heterogeneity in a population via individual-level covariates. The use of individual-level models for seasonal influenza dynamics is important since influenza is usually transmitted from person to person by direct contact. However, in most cases due to confidentiality, the information related on individual movement and contact behavior is very limited. When aggregated information is available, compartmental models are commonly used to model the spread of infectious disease (Corberán & Santonja, 2014).

In a surveillance setting, the number of infections and susceptibles allow a consideration of the size and duration of outbreaks. Ideally what is required at each time point are the number of infected and the number of susceptible individuals. However, the information about susceptibles is rarely available. Finkenstadt et al. (2002) reconstructed the number of susceptibles using data on cases and births in order to explain extinction and recurrence of epidemics observed in measles. In our case, the number of susceptibles for influenza is constructed over time through recursion equations.

As Finkenstadt et al. (2002) pointed out, the pattern of local extinction is determined by three factors: (i) size and recruitment into the susceptible class (ii) rate of contact of susceptible and infected individuals within the community and (iii) rate of contact by local susceptibles with infecteds from other populations. Point (iii), is related to the migration or influx of infection, which we refer to as an influx process (Morton & Finkenstadt, 2005). This is subject to seasonal variation due to heterogeneities of social mixing and mobility of individuals. It has been noted that the seasonal variation in transmissibility is very important as it is capable of producing longer term oscillations in the dynamics of epidemics (Held & Paul, 2012).

In this paper, we present a compartmental model in discrete time to describe the pattern of weekly reported data for seasonal influenza in the province of Manitoba, Canada. Our aim is to capture both epidemic and endemic dynamics through modelling the disease transmission. The transmission parameter is allowed to vary over time, and we assume that the disease can only be reintroduced after extinction if there is a contact with infected individuals from other host populations. Previously, the transmission parameter was modeled based on weekly or bi-weekly effects (Corberań-Vallet et al., 2014; Finkenstadt et al., 2000, 2002, 2005; Held & Paul, 2012). Unlike that setting, we assume that the exact timing and duration (onset, peak, and end) of influenza activity may not fall in the same week from year to year, but within the same month approximately. In order to capture the variability of influenza activities from one season to the next, we model the seasonality with a 4-week period effect, i.e. approximate monthly effect, over years. Our model formulation takes into account a decomposition of disease incidence into an epidemic component following autoregressive dynamics on the number of cases at the previous time periods, and an influx process. As the time and the state variables are assumed to be discrete, this model is considered to be a discrete time, stochastic, Susceptible-Infectious-Recovered-Susceptible (DT-SIRS) model, where weekly counts of disease are assumed to follow a Poisson distribution. DT-SIRS models have some advantages over the continuous-time formulations. DT-SIRS models are often simpler to formulate and more readily understood than continuous-time

models, especially on those occasions where the data have been recorded over predetermined discrete time intervals (Corberań-Vallet et al., 2014).

Our framework is different from that used by Yaari et al. (2013), who considered an age-of-infection setting under a frequentist approach. Here we analyze the model from a Bayesian viewpoint. In addition, the discrete formulation of the model better adjusts to the available data (weekly counts) and simplifies the Bayesian analysis. The Bayesian approach allows us to consider a probability distribution for the observed counts of disease as well as for the parameters of the model.

This paper is organized as follows. In Section 2, we present the data in detail. Next, we describe our DT-SIRS model. Section 3 shows the results obtained from the analysis of the influenza data from the central Canadian province of Manitoba between 2012 and 2015. Finally, we conclude the paper with a general discussion of the proposed model and provide directions for future research.

2. Methods

2.1. Data description

A number of public and private health agencies collect and distribute surveillance reports on predominant epidemics or other diseases of interest. FluWatch is Canada's national surveillance system that monitors the spread of flu and other flu-like illnesses on an ongoing basis. This program consists of a management and surveillance information reporting system for influenza and public health emergencies to significantly improve disease monitoring and early warning through network reporting (PHAC, 2019). In this study, we focus on influenza type A incidence data from Public Health in the Province of Manitoba. We consider weekly influenza incidence data reported from the 1st week of January 2012 to the 52nd week of December 2015, which is equivalent to 208 weeks. This data set shows an interesting structure in which years 2012 and 2015 are considered as extreme cases (low and high respectively).

For illustration, Fig. 1 shows the time trend of the data, which presents a clear seasonal pattern. The plot represents the time series of weekly counts for influenza data. These influenza counts show yearly outbreaks of different severity (height of peak) during the winters. As we observe in Fig. 1, there is only one peak per epidemic year which is located between week 49, and week 12 of the following year, most often between week 52, and week 3 of the following year. As it can be seen, flu incidence reaches the highest rates during winter. From spring the trend is descending, with very few cases during the summer and part of fall. In the end of fall, we can observe an upward trend that continues during the winter to complete the seasonal cycle. Fig. 2 shows that the epidemic activities gradually decrease after week 12 in each epidemic year. Lower values are observed from week 18 to week 48 and there is almost no epidemic activity in the summer. Table 1 gives informative summary statistics of the data where the highest weekly reported value is 267 and the smallest value is zero. In Fig. 2 we visualize the weeks with higher and lower incidences.

As the plot of the time series shows, given a population of the entire province, most of the individuals are susceptible during epidemic and endemic periods. The model is defined considering that the size of the susceptible compartment is large

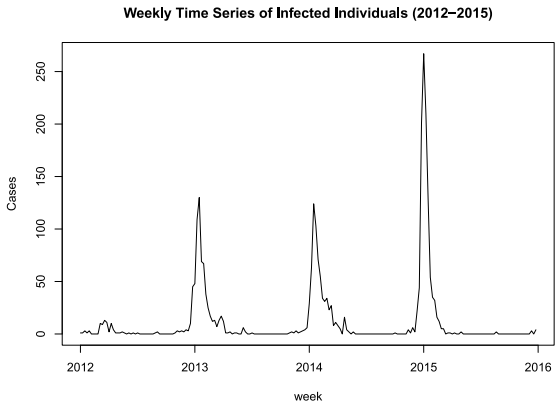

Fig. 1. Number of weekly observed incidences of influenza in the province of Manitoba from week 01, 2012 to week 52, 2015

Table 1
Summary of weekly cases (weekly reported data, January 2012 to December 2015).

Min	25th Per	**Median**	**Mean**	75th Per	90th Per	95th Per	**Max**
0	0	1	11.9	5	32.6	65.6	267

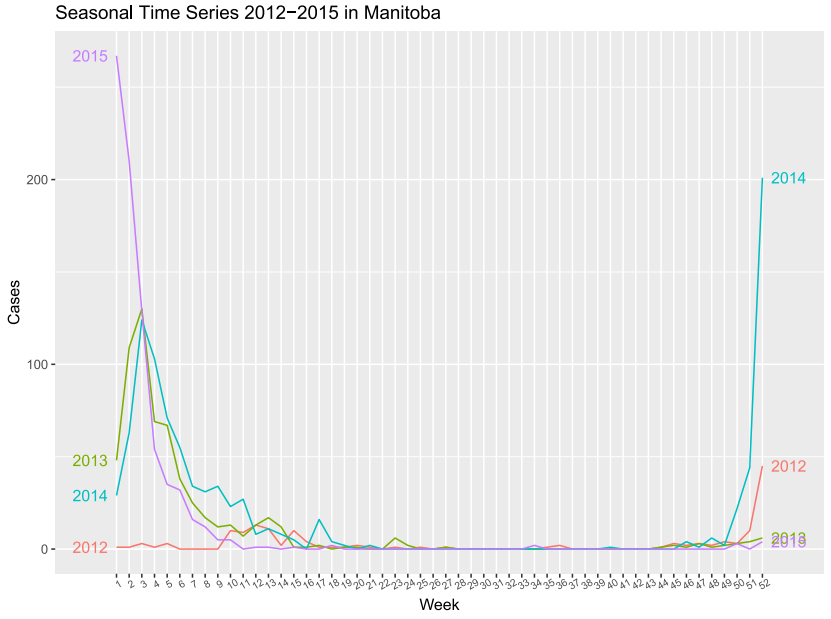

Fig. 2. Seasonal plot of the number of weekly observed incidences of influenza in the province of Manitoba (January 2012 to December 2015).

enough so that the mixing of susceptible individuals in the population is homogeneous. In order to provide a realistic description of the data, as it was explained in the introduction, we propose a DT-SIRS model in discrete time.

2.2. DT-SIRS model

In this section, we present our Bayesian stochastic SIRS model in discrete time with a focus on describing seasonal influenza dynamics. We consider at any given time the number of susceptible, infected, and recovered individuals from the population, corresponding to SIR models. The infectious individuals spread the disease during a determined length of time, then after they recover they acquire a partial immunity. Based on the specific property of population growth, the spread rules of infectious diseases, and the related social factors, we propose a model that reflects the dynamical behavior of infectious diseases over time by including the number of cases in the previous time period as a potential explanatory variable for the disease incidence. In this regard, to explain the spread of disease over time, we incorporate an autoregression on the number of cases at the previous time points $t-1$ and $t-2$. The inclusion of previous cases allows for temporal dependence beyond seasonal patterns within a region of interest. However, such a model will not be able to account for outbreaks due to possible infected visitors. With a DT-SIRS model, we consider the dynamics of the epidemic in a population which consists of resident and visitor (influx process) individuals.

Let I_t represent the number of newly infected individuals in week t and S_t and R_t denote the number of susceptible and recovered individuals in week t ($t = 1, \ldots, n$) respectively. Given the numbers of infected, I_{t-1}, and susceptibles, S_{t-1}, in week $t-1$, the number of infected individuals, I_t, in week t is assumed to be a random variable having a Poisson distribution with mean λ_t:

$$\begin{aligned} I_t|\lambda_t &\sim Pois(\lambda_t) \\ \lambda_t &= S_{t-1}(I_{t-1} + I_{t-2} + P_{t-1})^{\alpha}\exp\{r_t\} \\ P_t &\sim Pois(\lambda_p), \end{aligned} \tag{1}$$

where P_t is an another key stochastic process that indicates the influx of infected individuals from other places (the number of epidemic imports) at time t, and r_t is the transmission rate, discussed in more detail below, while α is a mixing parameter, modifing the influence of previous infected and the influx of infected individuals in the model. The parameter α allows for the nonlinearities in contact rates that may arise due to population substructuring or other forms of heterogenous mixing. If homogenous mixing in the population is assumed, this means that everyone interacts with equal probability with everyone else; which discards possible heterogeneities arising from age, space, or behavioral aspects as expected here.

Under the discrete-time model, the number of individuals in each compartment is considered at discrete time steps. Here, the number of susceptible, S_t and recovered, R_t, individuals at week t are updated through the following recursion equations:

$$\begin{aligned} S_t &= S_{t-1} - I_{t-1} + \psi R_{t-1} \\ R_t &= (1-\psi)R_{t-1} + \eta I_{t-1}, \end{aligned}$$

where ψ indicates the proportion of recovered individuals who lose their immunity and become susceptible again per unit of time and $1/\psi$ is the average time an individual remains immune against influenza re-infection; η is the proportion of infectious individuals that recover from a disease per unit of time (weeks) and $1/\eta$ is the average time to recover from disease (Keeling & Rohani, 2008). The transmission rate, r_t, is allowed to vary over time. In order to capture the seasonal dependence, we consider different models based on a variety of expressions for r_t.

Model 0 (M_0), in this paper, models the expressions of r_t that have been considered in the literature and are being employed as a benchmark for comparison purposes. The first version used considers fixed weekly seasonality effect over years (Corberán & Santonja, 2014; Finkenstadt & Grenfell, 2000; Morton & Finkenstadt, 2005). The other version uses sine-cosine waves to model the seasonal variation (Corberán & Santonja, 2014; Paul et al., 2008). Both of these versions of r_t are considered here.

- In **Model 0a** (M_{0a}), the transmission rate, r_t, varies over time with a 1-year period (52 weeks) where the weekly effect, φ_w, is the same for each week over years, so there are 52 fixed values $\varphi_1, \ldots, \varphi_{52}$, as described in the following expressions:

$$r_t = \varphi_{w(t)} \tag{2}$$

where,

$$w(t) = \begin{cases} p(t), & \text{if } p(t) \neq 0 \\ 52, & \text{otherwise.} \end{cases}$$

and

$$p(t) = t \mod 52.$$

- A further reduction in the number of parameters could, for example, be achieved by fitting **Model 0b** (M_{0b}) where the transmission rate, r_t, is represented as a sinusoidal curve of the form

$$r_t = \sum_{h=1}^{H} \left[\gamma_{2h} sin\left(\frac{2\pi}{52} ht\right) + \gamma_{2h-1} cos\left(\frac{2\pi}{52} ht\right) \right], \tag{3}$$

where H represents the number of sine-cosine waves needed to capture the seasonal variation in the disease transmission.

In **Model 1** (M_1), r_t is modeled using a fixed 4-week period effect over years. In this case, there are 13 temporal fixed effect parameters, $\varphi_1, \ldots, \varphi_{13}$, representing 4-week period effects, and

$$r_t = \varphi_{m(t)}, \tag{4}$$

where $1 \leqslant m(t) \leqslant 13$ and is defined as follows:

$$m(t) = \begin{cases} q(t) \mod 13, \\ 13, \quad \text{if } q(t) \mod 13 = 0 \end{cases}$$

with

$$q(t) = \begin{cases} floor\left(\frac{t}{4}\right) + 1, & \text{if } mod\left(\frac{t}{4}\right) \neq 0 \\ floor\left(\frac{t}{4}\right), & \text{if } mod\left(\frac{t}{4}\right) = 0, \end{cases}$$

where φ_m represents the effect over 4 consecutive weeks of the year.

Model 2 (M_2) considers r_t with different 4-week period effects over years:

$$r_t = \varphi_{q(t)}, \tag{5}$$

where

$$q(t) = \begin{cases} floor\left(\frac{t}{4}\right) + 1, & \text{if } mod\left(\frac{t}{4}\right) \neq 0 \\ floor\left(\frac{t}{4}\right), & \text{if } mod\left(\frac{t}{4}\right) = 0. \end{cases}$$

Model 3 (M_3) defines r_t using a 4-week period effect with a multiplicative year effect:

$$r_t = \varphi_{m(t)}\delta_{y(t)}, \tag{6}$$

where $\varphi_{m(t)}$ was given in (4), $\delta_{y(t)}$ denotes a year-indicator and $y(t)$ is defined as:

$$y(t) = \begin{cases} floor\left(\frac{t}{52}\right) + 1, & \text{if } mod\left(\frac{t}{52}\right) \neq 0 \\ floor\left(\frac{t}{52}\right), & \text{if } mod\left(\frac{t}{52}\right) = 0. \end{cases}$$

In both models M_0 and M_1, the transmission rate is allowed to vary through time with a 1-year period to accommodate the seasonal cycle. However, M_1 provides a considerable reduction in the number of parameters. All models M_1, M_2 and M_3 capture the 4-week period effect, with M_1 utilizing the same 4-week period effect over years; M_2 indicates a different 4-week period effect across each year and M_3 captures the same 4-week period effect over years. However, for models M_2 and M_3, the number of parameters increases or decreases with the size of the data set and is scaled by a year effect.

3. Results and discussion

3.1. Model selection and parameter estimation

The data set indicates that the seasonal influenza shows different intensities over years. Years 2012 and 2015 can be considered as unusual years with extreme cases. Separate models were fitted for three targeted data sets: the first data set includes both the smallest and largest peaks of infection activity (Fig. 3), i.e data from 2012 to 2015, the second data set omits the smallest peak (Fig. 4), i.e. data from 2013 to 2015 and the third data set omits the largest peak of 2015 (Fig. 5), i.e. data from 2012 to 2014, and. Among those three proposed versions of the model, M_2 shows a good fit compared to M_1 and M_3 respectively. Goodness-of-fit measures are discussed later in this section.

Like many other statistical models, the resultant posterior distribution of the model parameters is not analytically tractable. Hence, posterior sampling was carried out using Markov chain Monte Carlo(MCMC) simulation with the free statistical software WinBUGS (Lunn et al., 2000). The MCMC methods are used to sample from the posterior distribution and to estimate parameters. We specify prior distributions for all model parameters α, φ, δ, ψ, γ and λ_p, and values for their corresponding hyper-parameters. Priors Normal distributions with moderate precision parameters are assigned to φ and γ, and uniform distributions to δ, α, ψ and η. We fixed a burn-in period of 250000 iterations to assess the convergence of MCMC chains. To

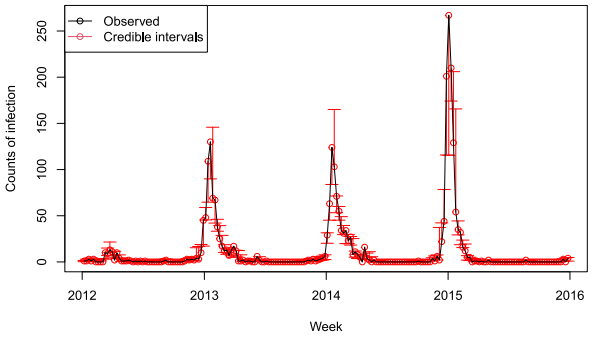

Fig. 3. Reported data of seasonal influenza (black) and 95% credible intervals (red) corresponding to the first week of January 2012 to the last week of December 2015. The credible intervals are the intervals in the domain of a posterior probability distribution of λ_t estimated from the data and obtained with the Bayesian stochastic DT-SIRS model where the transmission rate is modeled incorporating a different 4-week period effect over years (M_2).

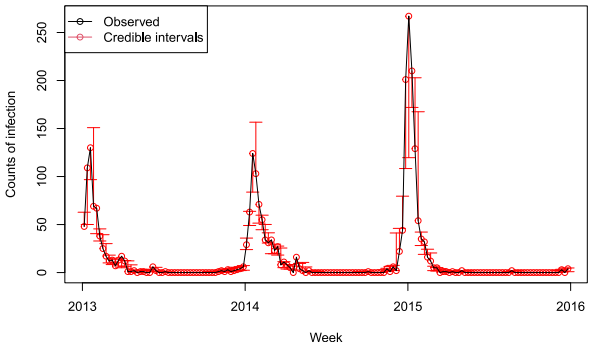

Fig. 4. Reported data of seasonal influenza (black) and 95% credible intervals (red) corresponding to the first week of January 2013 to the last week of December 2015. The credible intervals are the intervals in the domain of a posterior probability distribution of λ_t estimated from the data and obtained with the Bayesian stochastic DT-SIRS model where the transmission rate is modeled incorporating a different 4-week period effect over years (M_2).

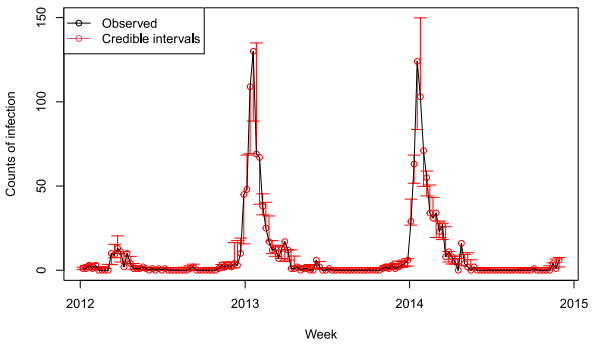

Fig. 5. Reported data of seasonal influenza (black) and 95% credible intervals (red) corresponding to the first week of January 2012 to the last week of November 2014. The credible intervals are the intervals in the domain of a posterior probability distribution of λ_t estimated from the data and obtained with the Bayesian stochastic DT-SIRS model where the transmission rate is modeled incorporating a different 4-week period effect over years (M_2).

reduce the correlation for the samples, we consider 1 posterior sample in 50 iterations after the burn-in-period until a posterior sample of size $n = 3000$ was obtained. Convergence of MCMC chains was assessed using the Gelman-Rubin statistics with jagsUI package in R software.

To compare the above mentioned models, we use the deviance information criterion (DIC), defined as $DIC = \bar{D}(\theta) + pD$, where $D(\theta)$ is the posterior mean of the deviance and $\theta = (\alpha, \varphi, \delta, \psi, \gamma, \lambda_p)$ denotes the collection of parameters in the model (Spiegelhalter et al., 2002). The penalty term pD is the effective number of model parameters, defined by $pD = \bar{D}(\theta) - D(\bar{\theta})$, where $\bar{\theta} = E(\theta|Y)$ is the posterior mean of θ. Models with lower DIC values are preferred as they achieve a more optimal combination of fit and parsimony (Gelman et al., 2013) (see Table 2 below). Additional measures of goodness of fit are based on comparing reported and predicted quantities as well as numerical data summaries. Those measures are root mean squared error (RMSE) (see Table 2) and graphical plots (see Figs. 3–5). Figs. 3–5 represent the reported influenza data together with the posterior credible intervals under the Poisson model, displayed under the model M_2, compared to observed data. Similar fits were obtained with M_1 and M_3 (plots are provided in the appendix).

A comparison of DIC for models M_{0a} and M_2 suggests that the two models are close in terms of overall fit but are superior to models M_1 and M_3. In addition, the comparison of RMSE shows that model M_2 is better than model M_{0a}. Practically, we might prefer reporting on model M_2 since its DIC is slightly greater than the more complex model M_{0a}. Both Figs. 3 and 4 show that the model M_2, which represents the model with r_t modeled using different 4-week period effects over years, fits the data well. Table 2 indicates that M_2 leads to an improved goodness of fit as judged by a lower DIC, among the proposed models, and lowest RMSE.

One of the special interests in infectious disease suveillance system is “short-term”forecast, especially one-step-ahead forecast. In this regard, one-step-ahead and two-step-ahead forecast performed well as discussed in this section. There are a number of different approaches to using Bayesian time series models to perform forecasting. One approach might be to fit a model, and use those posterior distributions to forecast as a secondary step. A more streamlined approach is to do this within the JAGS code itself. In this setting, we consider the latter option by taking advantage of the fact that JAGS allows us to include “NAs”(missing values) in the response variable (but never in the predictors). We omit the technical details here, but assume only that with obvious modifications, the model can be written down for data $I_1, I_2, \ldots, I_t$, NA where the count I_{t+1} is considered as a missing value in the case of forecasting one-step-ahead or $I_1, I_2, \ldots, I_t$, NA, NA where the counts I_{t+1} and I_{t+2} are considered

Table 2
DIC, pD and RMSE for different models: model M_0 where r_t is modeled using a fixed week seasonality effect over years; model M_1 where r_t is modeled using a fixed 4-week period effect over years; model M_2 where r_t is modeled using different 4-week period effects over years and model M_3 where r_t is modeled using 4-week period effect and multiplicative year effect. RMSE-Mean and RMSE-median are root mean squared error calculated using mean and median of 1000 replications of the posterior predictive distribution respectively.

Model	DIC	pD	RMSE-Mean	RMSE-Median
M_{0a}	623.96	76.2	4.98	7.62
M_{0b}	796.22	54.5	37.85	36.38
M_1	776.10	57.4	15.18	12.53
M_2	686.55	51.7	0.02	1.75
M_2 without I_{t-2}	1084.67	190.7	135.60	127.09
M_2 without P_t	1321.52	40.6	169.06	171.54
M_3	763.56	58.6	14.12	11.15

Table 3
Mean absolute error (MAE) and Mean absolute percentage error (MAPE) for a one-step ahead forecast.

Model	MAE	MAPE
M_{0a}	44.3	1.3
M_1	38.7	1.1
M_2	42.1	1.3
M_3	42.1	1.0

Table 4
Mean absolute error (MAE) and Mean absolute percentage error (MAPE) for a two-step ahead forecast.

Model	MAE	MAPE
M_{0a}	67.8	3.1
M_1	50.1	1.4
M_2	50.8	1.6
M_3	42.7	1.0

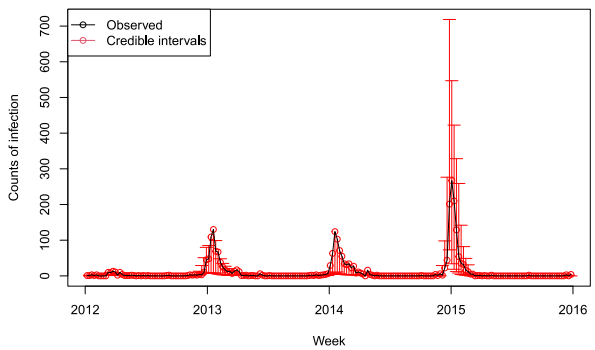

Fig. 6. Reported data of seasonal influenza (black) and 95% credible intervals (red) corresponding to the first week of January 2012 to the last week of December 2015. The credible intervals are the intervals in the domain of a posterior probability distribution of λ_t without the I_{t-2} term and estimated from the data and obtained with the Bayesian stochastic DT-SIRS model where the transmission rate is modeled by means of different 4-week period effect over years (M_2).

as missing values in the case of forecasting two-step-ahead. The n-step predictions were obtained by sequentially repeating the process up to time $t + n$. Tables 3 and 4 present the measures of prediction accuracy of one-step and two-step forecasting under the proposed models.

The results from Table 3 indicate that the all models have similar performances in terms of forecasting and Table 4 shows that the three models $M1$, M_2 and M_3 forecast better than M_0. Based on this forecast result and the one for measures of model complexity and fit in Table 2, the model M_2 is preferred.

Two measures of prediction accuracy of one-step and two-step forecasting used are the following.

- Mean absolute error (MAE):

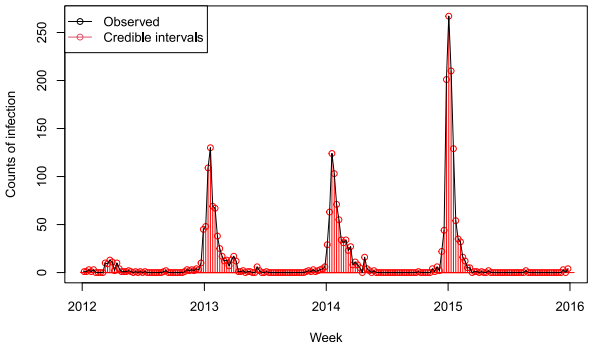

Fig. 7. Reported data of seasonal influenza (black) and 95% credible intervals (red) corresponding to the first week of January 2012 to the last week of December 2015. The credible intervals are the intervals in the domain of a posterior probability distribution of λ_t and estimated from the data and obtained with the Bayesian stochastic DT-SIRS model where the transmission rate is modeled by means of different 4-week period effect over years (M_2) with P_t set identically to zero.

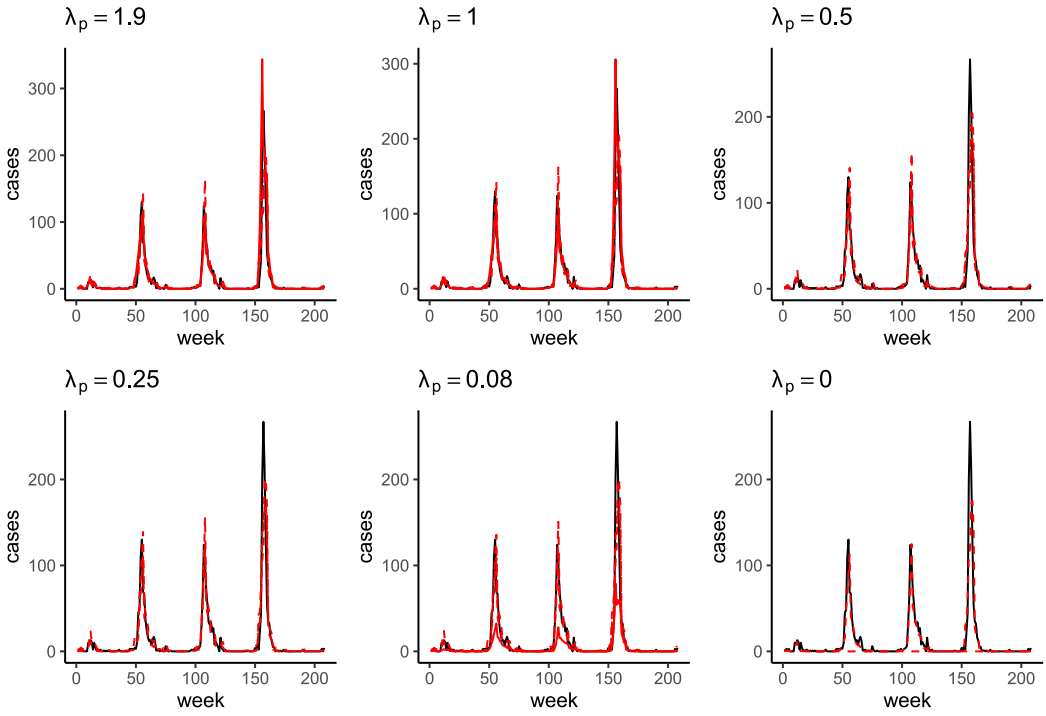

Fig. 8. The impact of influx process on the infection dynamics when the λ_p changes.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|$$

- Mean Absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}||\frac{A_t - F_t}{A_t},$$

where A_t is the actual reported value, F_t is the forecast value, and n is the number of steps.

Table 5
The impact of influx process with respect to infection dynamics. The mean of the seasonal peaks is the mean value of the peak count of infected over 2013, 2014 and 2015. The third column provides the average % change in peak values over these years where peak values under various values of λ_p are obtained as mean values from 1000 draws of estimates models based on M_2 with λ_p fixed.

Rate of ($\hat{\lambda}_p = 1.9$)	Mean of seasonal peaks(MSP = 188)	% MSP change
$\lambda_p = 1$	169	10
$\lambda_p = 0.5$	140	25
$\lambda_p = 0.25$	101	46
$\lambda_p = 0.08$	47	75
$\lambda_p = 0$	0	100

The structure of the model and the availability of information allow us to forecast only two periods ahead. We do not provide a plot for the forecast because we are performing an iterative forecast (one-step and two-step). The MAPE is scale sensitive and should not be used when working with data that contains zeroes. Note that because "actual reported data "is in the denominator of the equation, the MAPE is undefined when actual reported is zero. Furthermore, when the actual value is not zero, but quite small, the MAPE will often take extreme values. The approximate approach used in this article to overcome this issue was to replace the zeroes in our observed data with a small number, 1 in this case.

3.2. Influence of autoregressive I_{t-2} and influx process P_t components

In Table 2, we show goodness of fit statistics for a number of models. The absence of I_{t-2} in the model indicates lack of fit and also confirms the importance of including the autoregressive of order 2 in the model. Fig. 6 shows how the model without I_{t-2} overestimates the year 2015.

Similarly, we see a poorer fit when we remove the influx process term P_t. The term P_t is a key component in the model. It indicates the important influence of immigration in terms of extinction and recurrence of the epidemics. Fig. 7 shows how the epidemic dies out after the first wave if P_t is set identically to zero. An influx process is needed here for disease propagation. This results show that the reintroduction of the disease through contact with an influx of infected individuals from other regions is an important stochastic component. Hence, the influx process P_t plays a fundamental role in determining the epidemic dynamics for the Province of Manitoba. If there was no influx at all, the disease would remain extinct after an epidemic outbreak.

Fig. 8 shows the mean values from 1000 draws of the estimated model based on M_2 with λ_p fixed at various values. An importatnt public health quantinty is the peak count of infected individuals and we consider differences in these peaks for estimates from model with various values of λ_p. In particular, we calculate the difference between the mean peak values over the 1000 simulations of the fitted model when λ_p takes values of 1, 0.5, 0.25, 0.08 and 0 for each of the years 2013, 2014 and 2015 as a percentage change from the corresponding values for $\lambda_p = \hat{\lambda}_p$, the posterior mean estimate, 1.9. Table 5 provides the mean values of these percentage changes over the three years.

4. Conclusion

In this article, we make a use of a DT-SIRS model including recruitment of the susceptible population through loss of immunity and from immigration to the region. We also consider a new way of modelling the transmission rate. In the model every individual in the population is assumed to belong in only one of the three classes (susceptible, infected, recovered) at a time. It is assumed that susceptible individuals move to the I (infected) class after contact with an infective. Infective individuals recover after a period of time and move to the R (recovered) class, whereas recovered individuals eventually lose their immunity and become susceptible once again. The exact timing and duration (onset, peak, and end) of influenza activity do not fall in the same week from year to year. We utilized 4-week period effect in order to model the seasonality. The model accommodates many parameters, which we estimate with considering a Bayesian framework.

In addition, this work was motivated by the need to model seasonal influenza dynamics during different periods in the province of Manitoba. Three different expressions for transmission rate were considered and then were compared with others from the literature. Our model shows a good fit to the data as it is able to capture different peaks and troughs. Even though there is limited information in terms of the number of susceptibles and recovered, the model is able to reproduce the behavior of transmission of the disease over time and is able to forecast for short time periods of the year with higher or lower intensity of the infection. We have demonstrated here that the simple models for transmission rates effectively capture the behaviour of the disease dynamics.

It is also very clear that the influx process is a critical component for propagating the disease over time. This is also evident in the current COVID-19 pandemic where mobility is critical in management of the disease in order to limit the influx of cases.

An extension of the proposed model into the spatial domain will be particularly useful, since geographical space can play a significant role in disease dynamics. In addition, since excess zeros can occur when the data are considered on a weekly frequency; our plan is to study a zero-inflated model for this kind of data.

Another very fruitful area for future research is the incorporation of covariates affecting disease transmission, for instance socio-economic, environmental and demographic factors which greatly improve description and prediction of infectious disease dynamics over time or across different areas.

For a spatio-temporal model, we would like to know whether there are some influential areas in terms of transmission dynamics of the disease, if the season starts at the same time every year across all areas, or if there are patterns in terms of spread (i.e. whether there are areas that always have earlier or late outbreaks).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Statistical Sciences Institute- Collaborative Research Teams (CANSSI-CRT) grants. Special thanks to Elizabeth Renouf for her invaluable discussion.

Appendix A. Model with same 4-week period effect over years (M_1)

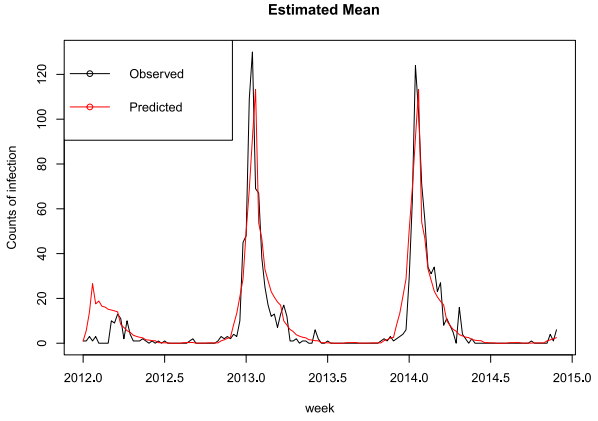

Fig. A.1. Mean of 1000 simulations of fitted values (red) and reported counts (black).

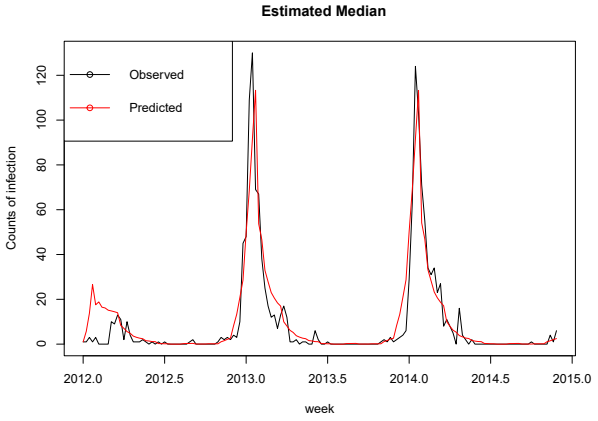

Fig. A.2. Median of 1000 simulations of fitted values (red) and reported counts (black).

Appendix B. Model with 4-week period effect and multiplicative year effect (M_3)

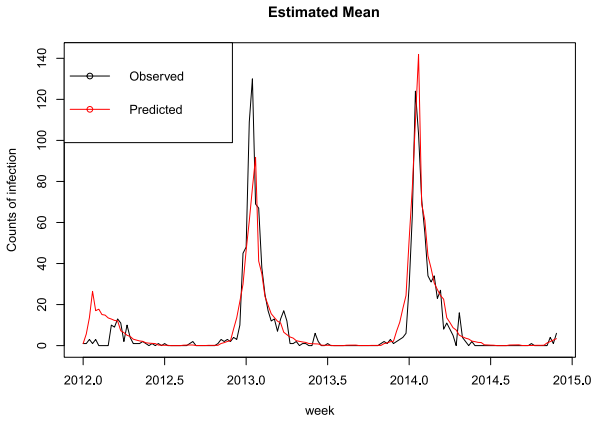

Fig. B.1. Mean of 1000 simulations of fitted values (red) and reported counts (black).

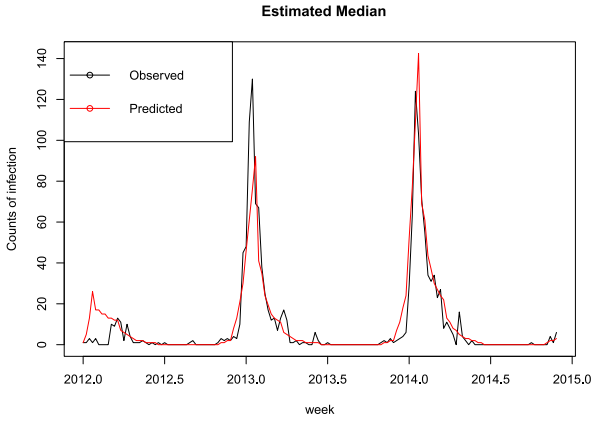

Fig. B.2. Median of 1000 simulations of fitted values (red) and reported counts (black).

References

Corberán, A. V., & Santonja, F. J. (2014). A Bayesian SIRS model for the analysis of respiratory syncytial virus in the region of Valencia, Spain. *Biometrical Journal, 56*(5), 808–818.
Deeth, L.,E., & Deardon, R. (2016). Spatial data aggregation for spatio-temporal individual-level models of infectious disease transmission. *Spatial and Spatio-temporal epidemiology, 17*, 95–104.
Finkenstadt, B. F., Bjornstad, O. N., & Grenfell, B. T. (2002). A stochastic model for extinction and recurrence of epidemics: Estimation and inference for measles outbreaks. *Biostatistics, 3*(4), 493–510.
Finkenstadt, B. F., & Grenfell, B. T. (2000). Time series modelling of childhood diseases: A dynamical systems approach. *Applied Statistics, 49*(2), 187–205.
Finkenstadt, B. F., Morton, A., & Rand, D. A. (2005). Modelling antigenic drift in weekly flu incidence. *Statistics in Medecine, 24*, 3447–3461.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
Held, L., & Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal, 54*(6), 824–843.
Keeling, M. J., & Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Chap.2.
Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical modelling in spatial epidemiology* (3rd ed..).
Lawson, A. B., & Song, H. R. (2010). Bayesian hierarchical modeling of the dynamics of spatio-temporal influenza season outbreaks. *Spatial and spatio-temporal epidemiology, 1*, 187–195.
Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a bayasian modelling framework: Concepts, structures, and extensibility. *Statistics and Computing, 10*, 325–337.
Morton, A., & Finkenstadt, B. F. (2005). Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, 54*(3), 575–594.
Paul, M., Held, L., & Toschke, A. M. (2008). Multivariate Modelling of infectious disease surveillance data. *Statistics in Medecine, 27*, 6250–6267.
Public Health Agency of Canada (PHA). (2012). Seasonal influenza - infection prevention and control guidance for management in home care settings. https://www.canada.ca/en/public-health/services/infectious-diseases/nosocomial-occupational-infections/seasonal-influenza-infection-prevention-control-guidance-management-home-care-settings. Available at:.
Public Health Agency of Canada (PHA). (2019). Overview of influenza monitoring in Canada. https://www.canada.ca/en/public-health/services/diseases/flu-influenza/influenza-surveillance/about-fluwatch.html. Available at:.

Public Health Agency of Canada (PHA). (2021). Weekly influenza reports in Canada. https://www.canada.ca/en/public-health/services/diseases/flu-influenza/influenza-surveillance/weekly-influenza-reports.html. Available at:.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology), 64*(4), 583–639.

Wi. (2021). Worldometers.info (WI). https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/.

Yaari, R., Katriel, G., Huppert, A., & Stone, L. (2013). Modelling seasonal influenza: The role of weather and punctuated antigenic drift. *Journal of The Royal Society Interface, 10*, Article 20130298.