WILEY | Hindawi

Research Article

Modeling Chickenpox Dynamics with a Discrete Time Bayesian Stochastic Compartmental Model

A. Corberán-Vallet,[1] F. J. Santonja (iD),[1] M. Jornet-Sanz,[1] and R.-J. Villanueva (iD)[2]

[1]*Department of Statistics and Operational Research, Universitat de València, Valencia, Spain*
[2]*Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain*

Correspondence should be addressed to F. J. Santonja; francisco.santonja@uv.es

We present a Bayesian stochastic susceptible-exposed-infectious-recovered model in discrete time to understand chickenpox transmission in the Valencian Community, Spain. During the last decades, different strategies have been introduced in the routine immunization program in order to reduce the impact of this disease, which remains a public health's great concern. Under this scenario, a model capable of explaining closely the dynamics of chickenpox under the different vaccination strategies is of utter importance to assess their effectiveness. The proposed model takes into account both heterogeneous mixing of individuals in the population and the inherent stochasticity in the transmission of the disease. As shown in a comparative study, these assumptions are fundamental to describe properly the evolution of the disease. The Bayesian analysis of the model allows us to calculate the posterior distribution of the model parameters and the posterior predictive distribution of chickenpox incidence, which facilitates the computation of point forecasts and prediction intervals.

1. Introduction

Chickenpox is a highly contagious disease caused by the varicella-zoster virus. Its main symptoms are a blister-like rash, itching, tiredness, and fever. It is a rarely fatal disease that mainly affects children younger than ten years of age, although older children and adults can also get it. Most of the people suffer the disease during childhood and reinfection is very strange [1]. Because of the contagious behavior of the disease, which spreads easily through the coughs and sneezes of an infected person or by touching the virus particles that come from chickenpox blisters, the best modeling approach would be obtained by taking into account individual movement and contact behavior. However, information at individual level is scarcely ever available and models based on aggregated data are usually considered.

Compartmental models are commonly used in epidemiology to understand the underlying mechanisms that influence disease transmission when aggregated data are available. These models divide the population being studied into different compartments according to the disease status and describe the evolution of the disease through changes in the number of individuals in each compartment. There may be susceptible, exposed, infectious, recovered, and immune individuals depending on the nature of the disease. Compartmental models are usually written using ordinary differential equations or difference equations. In addition, they are normally defined assuming that all the individuals in the population are equally likely to contact any other individual in that population [2, 3]. Models based on difference equations have been considered to describe the progression of chickenpox disease (see, e.g, [4, 5]). However, contact patterns in real populations are indeed heterogeneous. Therefore, models assuming homogeneous mixing should be replaced by models incorporating stochastic effects [2]. Stochastic models are able to accommodate the stochasticity inherent in the transmission of infectious diseases and provide an improved and more natural description of disease dynamics [6–8]. In addition, stochastic models can be analysed from a Bayesian viewpoint (see, e.g., [9]).

In the case of chickenpox, symptoms appear about two weeks (from ten to twenty-one days) after exposure to a

contagious person. A person with chickenpox can spread the disease from one to two days before they get the rash until all the blisters have fully crusted over. The condition usually resolves by itself within a week and the infected individual usually acquires permanent immunity [1]. Therefore, a susceptible-exposed-infectious-recovered (SEIR) model seems an appropriate model to describe chickenpox dynamics.

In this paper, we present a Bayesian stochastic SEIR model in discrete time to understand chickenpox dynamics in the Valencian Community, Spain. The proposed compartmental model takes into account both heterogeneous mixing of individuals in the population and the inherent stochasticity in the transmission of the disease. In addition, the discrete formulation of the model better adjusts to the available data (weekly counts) and simplifies its Bayesian analysis. The analysis of the model from a Bayesian viewpoint allows us to consider a probability distribution for the observed counts of disease as well as for the parameters of the model. The posterior distribution for these parameters and the predictive distribution for future observations can then be straightforwardly obtained. Note that point forecasts and prediction intervals can be of utter importance to implement effective public health measures to reduce the burden of the disease.

The remainder of this paper is organized as follows. In Section 2 we present the available chickenpox data. Section 3 describes the SEIR model and shows its Bayesian analysis. In Section 4 we show the results obtained in the analysis of the data. Finally, we conclude with a general discussion of the proposed model and provide directions for further research.

2. Data

In this paper we analyse weekly cases of chickenpox reported by the health sentinel network in the Valencian Community, Spain, for the period 2007–2012 (http://www.sp.san.gva.es). It is important to point out that there were no changes in the vaccination strategy during the period analysed. Starting in 2006, the chickenpox vaccine was introduced in the routine immunization schedule of the Valencian Community and, during the period 2007–2012, 10–14 year-old susceptible children were vaccinated. Because varicella vaccine was available in pharmacies, younger children could also be vaccinated against varicella beyond official guidelines.

For 2007–2012, an average of 19,000 cases of chickenpox have been reported per year in the region of Valencia, representing an average incidence rate of around 375 cases per 100,000 inhabitants. Figure 3 shows the time plot of the series, which presents a clear seasonal pattern. As can be seen, chickenpox incidence reaches the highest rates during spring. From June the trend is descending, with very few cases during the summer and part of autumn. In the end of autumn, we can observe an upward trend that continues during the winter to complete the seasonal cycle.

Chickenpox is a mandatory notifiable disease in the Valencian Community. On the other hand, population coverage by the Spanish Public Health System is almost universal (99.5%) and even persons not covered by it can also be attended to in public centers. Therefore, underreported chickenpox cases are assumed to be insignificant.

3. Model

Although it can vary from person to person, an individual takes around two weeks to show symptoms after being infected and, as we mentioned before, a person with chickenpox can spread the disease from one to two days before the rash appears until all the blisters have crusted over (usually five to six days after the start of the rash). Hence, in order to describe as accurately as possible the different stages of chickenpox while keeping a relatively simple model, we make the following assumptions: once an individual comes into contact with a person with chickenpox, there is an incubation period of two weeks and then the individual becomes infectious for a week. Even though there is a proportion of infected individuals that, after being recovered, become susceptible again, that proportion is so small that it will not be considered here. Moreover, reinfected individuals do not transmit the disease. Consequently, chickenpox dynamics can be explained via a compartmental model that divides the population into susceptible, exposed, infectious, and recovered individuals. In addition, this model should focus on the first movement, that is, the transition from the susceptible compartment to the exposed one. Note that the following transitions (to the infectious compartment and finally to the recovered one) come naturally due to the nature of the disease.

Let E_t^1 be the number of people at the first week of their incubation period. Because the data reported by the health sentinel network refer to the number of infectious people at each week, represented by I_t, it is necessary to transform the available data into the time-series $\{E_t^1\}$. This can be easily done taking into account the fact that I_{t+2} was E_t^1. At the first level of the model hierarchy, we can assume that E_t^1 follows a Binomial distribution:

$$E_t^1 \sim \mathrm{Bi}\left(S_{t-1}, p_t\right), \tag{1}$$

where S_{t-1} is the number of susceptible individuals at week $t-1$ and p_t is the probability of getting the virus at week t. In order to take into account the transmissible nature of the disease, we assume that the probability of infection depends on the number of infectious individuals at the previous week. Besides, this probability must lie in the intervals 0-1. In order to avoid the problem of estimates of p_t falling outside the 0-1 limits, we model logit(p_t) instead, which lies in the interval $(-\infty, \infty)$. In particular, we assume that

$$\mathrm{logit}\left(p_t\right) := \log\left(\frac{p_t}{1 - p_t}\right) = \alpha \log\left(I_{t-1}\right) + r_t, \tag{2}$$

which is equivalent to

$$p_t = \frac{I_{t-1}^{\alpha} \exp\{r_t\}}{1 + I_{t-1}^{\alpha} \exp\{r_t\}}, \tag{3}$$

where $\exp\{r_t\}$ is the transmission rate and the mixing parameter α in (2) allows for heterogeneous mixing of individuals
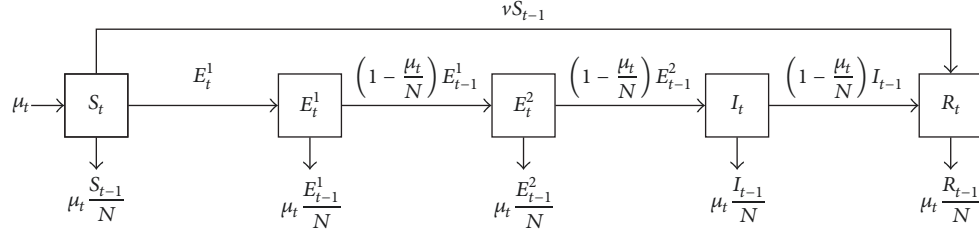
FIGURE 1: Flowchart of the proposed chickenpox model. Boxes represent compartments, horizontal arrows represent transitions between compartments, and downward arrows represent people leaving the system.

in the population [10]. Setting $\alpha = 1$ would correspond to the assumption of mass-action law. We assume here that α follows the uniform distribution in the interval $(0, 1)$. A key feature of our model formulation is that the transmission rate is allowed to vary stochastically over time. In particular, we model r_t as follows:

$$r_t = a_0 + \sum_{k=1}^{K} \left[a_{2k-1} \cos\left(\frac{2k\pi t}{52}\right) + a_{2k} \sin\left(\frac{2k\pi t}{52}\right) \right] \\ + \epsilon_t,$$
(4)

where K represents the number of sine-cosine waves needed to capture the seasonal variation in the disease transmission. Its value depends on the data under study, and so K has to be estimated together with the other parameters of the model. Because the seasonal pattern varies slightly from year to year, the random effect $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ allows for unspecified features of week t. The parameters $\{a_k\}$ are assumed to have zero mean Gaussian distributions with standard deviations σ_{a_k}. The uniform distribution in the interval $(0, 5)$ is assigned to all the standard deviations [11].

The number of individuals in each compartment is examined at discrete time steps through the following recursive equations:

$$S_t = S_{t-1} - E_t^1 - \nu S_{t-1} + \mu_t - \mu_t \frac{S_{t-1}}{N}$$

$$E_t^1 = E_t^1$$

$$E_t^2 = E_{t-1}^1 - \mu_t \frac{E_{t-1}^1}{N}$$
(5)

$$I_t = E_{t-1}^2 - \mu_t \frac{E_{t-1}^2}{N}$$

$$R_t = R_{t-1} + \nu S_{t-1} + I_{t-1} - \mu_t \frac{I_{t-1}}{N} - \mu_t \frac{R_{t-1}}{N},$$

where S_t is the number of susceptible individuals, E_t^1 is the number of individuals at the first week of their incubation period (which represents the data at hand, and so it is just set equal to its value), E_t^2 is the number of individuals at the second week of their incubation period, I_t is the number of infectious individuals, and R_t is the number of recovered (immune) individuals at week t. Although the common formulation of the SEIR model includes only one

compartment of exposed individuals, we divide here the compartment E_t into two classes, E_t^1 and E_t^2, in order to simplify the update of the recursive equations. By using this formulation, all the individuals in compartment E^1 at week $t-1$ move to compartment E^2 at week t and so on. So the only compartments where individuals can stay longer than one unit time are the susceptible and recovered compartments. The average proportion of individuals vaccinated each week is represented by ν. The parameter μ_t represents the number of births at week t and N the population size, which is assumed to be constant over time ($N = S_t + E_t^1 + E_t^2 + I_t + R_t, \forall t$). To keep the population size constant, μ_t individuals must leave the system each week, proportionally to the sizes of the compartments; that is, $\mu_t = \mu_t(S_{t-1}/N + E_{t-1}^1/N + E_{t-1}^2/N + I_{t-1}/N + R_{t-1}/N)$. The flowchart diagram for the model is described in Figure 1.

Using demographic data from the Spanish National Institute of Statistics (http://www.ine.es), we can estimate the average weekly number of births in the Valencian Community for years 2007–2012 as $1{,}002$, and so we set $\mu_t = 1{,}002 \, \forall t$ and $N = 5{,}061{,}245$. The annual vaccination coverage among children younger than 14 years of age in the Valencian Community is approximately 45% [12]. Thus, parameter ν can be estimated as 0.45 times the percentage of population aged less than 14 years divided by 52. Because the vaccination program started in 2006, we assume that children older than 14 years old are already protected.

Taking into account the following relationship between exposed and infectious individuals,

$$I_t = \left(1 - \frac{\mu}{N}\right) E_{t-1}^2 = \left(1 - \frac{\mu}{N}\right)^2 E_{t-2}^1,$$
(6)

the initial conditions for the recurrence equations can be computed as follows:

$$E_{t_{\text{ini}}}^1 = \frac{I_{t_{\text{ini}}+2}}{(1 - \mu/N)^2}$$

$$E_{t_{\text{ini}}}^2 = \frac{I_{t_{\text{ini}}+1}}{(1 - \mu/N)}$$
(7)

$$I_{t_{\text{ini}}} = I_{t_{\text{ini}}}$$

$$R_{t_{\text{ini}}} = \% \text{ people older than 14 years} \times N$$

$$S_{t_{\text{ini}}} = N - E_{t_{\text{ini}}}^1 - E_{t_{\text{ini}}}^2 - I_{t_{\text{ini}}} - R_{t_{\text{ini}}}.$$

TABLE 1: Posterior means and 95% credible intervals for the model parameters.

Parameter	Mean	Credible interval
α	0.59	$(0.48, 0.66)$
a_0	-11.18	$(-11.58, -10.57)$
a_1	0.42	$(0.38, 0.46)$
a_2	0.44	$(0.34, 0.58)$
a_3	-0.12	$(-0.17, -0.06)$
a_4	-0.34	$(-0.38, -0.29)$
a_5	-0.01	$(-0.05, 0.03)$
a_6	0.15	$(0.11, 0.19)$

4. Results

In this section we show the main results obtained in the analysis of notified cases of chickenpox in the Valencian Community for the period 2007–2012. Because we want to model the number of individuals at the first week of their incubation period (see (1)), we used chickenpox infections from week 4, 2007, to week 3, 2012, to calculate E_t^1 from week 2, 2007, to week 1, 2012, (see relationship in (6)). A total of $T = 260$ data were used to estimate the model. Data for the first 3 weeks of 2007 (I_1, I_2, and I_3) were used to compute the initial conditions for the recursive equations with $t_{\text{ini}} = 1$ (see (7)). Finally, data for the last 49 weeks (from week 4 to week 52, 2012) were left to evaluate the ability of our model to predict the dynamics of chickenpox infection.

As it happens with many statistical models, the posterior distribution of the model parameters is not analytically tractable and this makes necessary the use of efficient simulation algorithms to obtain a sample from that posterior distribution. All the analysis was implemented using the free statistical software R with the package $R2OpenBUGS$ that allows us to call $OpenBUGS$ within R to carry out posterior sampling using Markov chain Monte Carlo simulation [13]. The code for the proposed model can be seen in Pseudocode 1. To assess the convergence of MCMC chains, we fixed a burnin period of 50,000 iterations. We then kept one posterior sample in ten iterations until a set of 5,000 iterations was obtained.

Figure 4 shows the estimated transmission rate r_t together with its seasonal component and the estimated probability of getting the virus at each time point. In this case study, the seasonal component is given by the sum of three harmonic waves ($K = 3$). Higher values of K were not significant. As can be seen, seasonality has a strong influence on disease transmission. Nevertheless, it is important to incorporate random effects into the model to accommodate variability in the transmission rate from year to year. The posterior mean and 95% credible intervals for parameters α and $\{a_k\}$ are shown in Table 1. Note that the posterior mean of parameter α is 0.59. This value of α smaller than 1 corroborates that the mixing of individuals in the population is heterogeneous.

Figure 5 displays the fit of the model to the data originally reported by the health sentinel network. The estimated data for week $t + 2$, I_{t+2}, has been calculated as $S_{t-1} \cdot \hat{p}_t \cdot (1 - \mu/N)^2$ (see (1) and (6)). The fitting root mean squared error (RMSE) is 4.41.

Finally, in order to measure the out-of-sample forecast accuracy, we forecast the number of chickenpox infections for the last 49 weeks (from week 4 to week 52, 2012). These forecasts have been obtained by simulation as follows. Given a sample of size J (here $J = 5,000$) from the posterior distribution of the model parameters, we compute $I_{T+3+h}^{(j)} = E_{T+1+h}^{1(j)} \cdot (1 - \mu/N)^2$, where $T = 260$, $h = 1, 2, \ldots, 49$, and $j = 1, 2, \ldots, J$, and $E_{T+1+h}^{1(j)}$ is simulated from the following Binomial distribution:

$$\text{Bi}\left(S_{T+h}, p_{T+1+h}^{(j)}\right), \tag{8}$$

where

$$
\begin{aligned}
p_{T+1+h}^{(j)} &= \frac{\widehat{I}_{T+h}^{\alpha^{(j)}} \exp\left\{r_{T+1+h}^{(j)}\right\}}{1 + \widehat{I}_{T+h}^{\alpha^{(j)}} \exp\left\{r_{T+1+h}^{(j)}\right\}}, \\
r_{T+1+h}^{(j)} &= a_0^{(j)} + a_1^{(j)} \cos\left(\frac{2\pi\,(T+1+h)}{52}\right) \\
&\quad + a_2^{(j)} \sin\left(\frac{2\pi\,(T+1+h)}{52}\right) \\
&\quad + a_3^{(j)} \cos\left(\frac{4\pi\,(T+1+h)}{52}\right) \\
&\quad + a_4^{(j)} \sin\left(\frac{4\pi\,(T+1+h)}{52}\right) \\
&\quad + a_5^{(j)} \cos\left(\frac{6\pi\,(T+1+h)}{52}\right) \\
&\quad + a_6^{(j)} \sin\left(\frac{6\pi\,(T+1+h)}{52}\right).
\end{aligned}
\tag{9}
$$

These equations correspond to (3) and (4) for time $T + 1 + h$ computed at $I_{T+h} = \widehat{I}_{T+h}$, the point forecast, $(\alpha, a_0, a_1, a_2, a_3, a_4, a_5, a_6) = (\alpha^{(j)}, a_0^{(j)}, a_1^{(j)}, a_2^{(j)}, a_3^{(j)}, a_4^{(j)}, a_5^{(j)}, a_6^{(j)})$, and the jth values simulated from the posterior distribution of all the unknowns, assuming that the random effects in the transmission rate are equal to zero at time $T + 1 + h$. As explained in (4), the random effects, which are normally distributed with zero mean, capture unspecified features of each week, and so our predictions are based on the deterministic component of the transmission rate. Point forecasts and prediction intervals for each time point can be straightforwardly computed using the sample $\{I_{T+3+h}^{(j)}\}_{j=1}^{J}$. Figure 6 displays the forecasting results. As can be seen, the proposed model allows us to quite accurately predict the timing and magnitude of the chickenpox epidemic in 2012. The forecast RMSE is 64.24.

4.1. Comparison with Alternative Formulations of the SEIR Model. For comparative purposes, we present here the results obtained in the analysis of the data with four alternative formulations of the SEIR model. The first one corresponds to a stochastic formulation of the SEIR model in discrete time with a deterministic seasonal transmission rate. The

```
model{
E1[1] ~ dbin(p[1], Sini)
logit(p[1])  <- alpha1 * log(Iini) + r[1]
for(t in 2:T){
    E1[t] ~ dbin(p[t], S[t-1])
    logit(p[t]) <- alpha1 * log(I[t-1]) + r[t]
}
for(t in 1:T){
    r[t] <- a + a1*cos(2*3.1416*t/52) + a2*sin(2*3.1416*t/52)
              + a3*cos(4*3.1416*t/52) + a4*sin(4*3.1416*t/52)
              + a5*cos(6*3.1416*t/52) + a6*sin(6*3.1416*t/52) + epsilon[t]
    epsilon[t] ~ dnorm(0, tau.epsilon)
}
alpha1 ~ dunif(0, 1)
a ~ dnorm(0, tau.a)
tau.a <- 1 / pow(sigma.a, 2)
sigma.a ~ dunif(0, 5)
a1 ~ dnorm(0, tau.a1)
tau.a1 <- 1 / pow(sigma.a1, 2)
sigma.a1 ~ dunif(0, 5)
a2 ~ dnorm(0, tau.a2)
tau.a2 <- 1 / pow(sigma.a2, 2)
sigma.a2 ~ dunif(0, 5)
a3 ~ dnorm(0, tau.a3)
tau.a3 <- 1 / pow(sigma.a3, 2)
sigma.a3 ~ dunif(0, 5)
a4 ~ dnorm(0, tau.a4)
tau.a4 <- 1 / pow(sigma.a4, 2)
sigma.a4 ~ dunif(0, 5)
a5 ~ dnorm(0, tau.a5)
tau.a5 <- 1 / pow(sigma.a5, 2)
sigma.a5 ~ dunif(0, 5)
a6 ~ dnorm(0, tau.a6)
tau.a6 <- 1 / pow(sigma.a6, 2)
sigma.a6 ~ dunif(0, 5)
tau.epsilon <- 1 / pow(sigma.epsilon, 2)
sigma.epsilon ~ dunif(0, 5)
}
```

PSEUDOCODE 1: BUGS code for the Bayesian stochastic SEIR model in discrete time.

second formulation is the deterministic counterpart, that is, a deterministic model in discrete time with a seasonal transmission rate. The last two models correspond to deterministic continuous-time formulations.

Alternative 1 (stochastic model in discrete time with a deterministic seasonal transmission rate (SMDT-DTR)). This alternative formulation is similar to the one proposed in this paper, but the transmission rate is modeled in a deterministic way. The transmission rate is allowed to vary seasonally, but we assume that the seasonal pattern repeats over time; that is, there are no weekly random effects that account for overdispersion in the transmission rate. In particular, the SMDT-DTR assumes that

$$r_t = a_0 + \sum_{k=1}^{K} \left[a_{2k-1} \cos\left(\frac{2k\pi t}{52} \right) + a_{2k} \sin\left(\frac{2k\pi t}{52} \right) \right]. \quad (10)$$

We have implemented this model in a Bayesian framework using the package *R2OpenBUGS* in a similar way to our model.

Alternative 2 (deterministic model in discrete time with a seasonal transmission rate (DMDT)). This model is the deterministic counterpart of the stochastic discrete time model with a deterministic seasonal transmission rate. Within this model formulation, the number of people at the first week of their incubation period is estimated as

$$e_t^1 = S_{t-1} \frac{I_{t-1}^\alpha \exp\{r_t\}}{1 + I_{t-1}^\alpha \exp\{r_t\}}, \quad (11)$$

where r_t is modeled as in (10). So, instead of analysing directly E_t^1 and considering that E_t^1 is a random variable with its associated probability distribution, the model is defined in terms of an estimate, e_t^1, which is given by the mean of
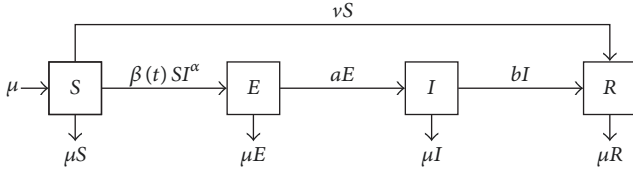
Figure 2: Flowchart of the continuous-time deterministic chickenpox model. Boxes represent compartments, horizontal arrows represent transitions between compartments, and downward arrows represent people leaving the system.

the Binomial distribution in (1). In this case, the recursive equations are given by

$$S_t = S_{t-1} - e_t^1 - \nu S_{t-1} + \mu_t - \mu_t \frac{S_{t-1}}{N}$$

$$E_t^2 = e_{t-1}^1 - \mu_t \frac{e_{t-1}^1}{N}$$

$$I_t = E_{t-1}^2 - \mu_t \frac{E_{t-1}^2}{N}$$

$$R_t = R_{t-1} + \nu S_{t-1} + I_{t-1} - \mu_t \frac{I_{t-1}}{N} - \mu_t \frac{R_{t-1}}{N}.$$

(12)

In order to fit this deterministic model, we minimized the sum of squared errors ($E_t^1 - e_t^1$) using the *optim* function in *R*.

Alternative 3 (deterministic continuous-time model with a seasonal transmission rate (DMCT)). As mention in the Introduction, deterministic compartmental models in continuous time are widely used to understand the dynamics of infectious diseases. A deterministic continuous-time SEIR model accounting for heterogeneous mixing patterns in the population can be formulated as follows:

$$\frac{dS}{dt} = -\beta(t) SI^\alpha - \nu S + \mu - \mu S$$

$$\frac{dE}{dt} = \beta(t) SI^\alpha - aE - \mu E$$

$$\frac{dI}{dt} = aE - bI - \mu I$$

$$\frac{dR}{dt} = \nu S + bI - \mu R,$$

(13)

where $\beta(t)$ is defined as $\gamma_0 + \gamma_1 \cos(2\pi t + \psi)$ to account for seasonality, μ is the proportion of births, and ν is the proportion of individuals vaccinated. The values of a and b are set taking into account the fact that the average time to show symptoms after being infected is 14 days and the average time to recover from chickenpox disease is 7 days. Therefore, in the continuous scenario, it is not necessary to distinguish between E^1 and E^2. Finally, we assume that $S + E + I + R = 1$, $\forall t$. The flowchart diagram for this model is described in Figure 2.
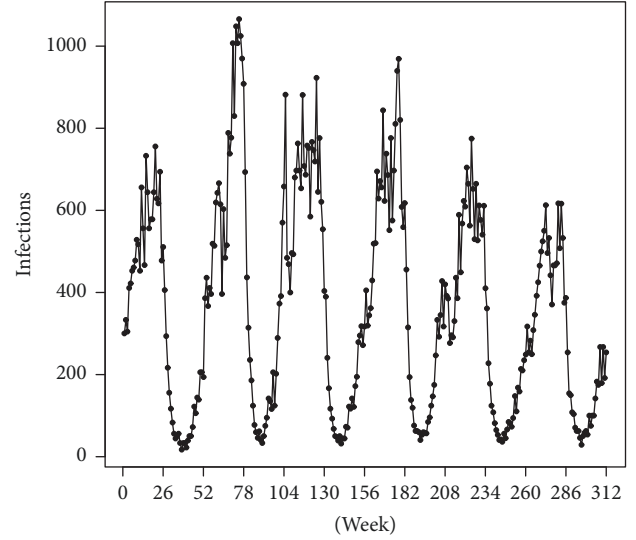

Figure 3: Weekly number of reported infections of chickenpox in the Valencian Community from 2007 to 2012.

Alternative 4 (deterministic continuous-time model with a seasonal transmission rate and assuming homogeneous mixing (DMCT-Homo)). This model is similar to the previous one, but it assumes that all the individuals in the population are equally likely to contact any other individual (homogeneous mixing), and so the model is formulated as in (13) setting $\alpha = 1$.

In order to estimate the parameters of these two previous continuous-time deterministic models, we have used the *R* package *deSolve* together with the *optim* function.

Table 2 shows the fitting and forecast RMSEs for the Bayesian stochastic SEIR model in discrete time proposed in this paper and the four alternative formulations previously described. As can be seen, the proposed model leads to an improved goodness of fit as judged by a lower RMSE. The comparison between the fitting error of the proposed model and that of the SMDT-DTR corroborates the idea that the incorporation of random effects when modeling the transmission rate is fundamental to explaining particular features of annual epidemics, which are different from year to year. The difference between the corresponding forecast errors is smaller due to the fact that the random effects in the transmission rate are set equal to zero when forecasting. Comparison between the fitting errors of the stochastic models and those associated with the deterministic approaches highlights the importance of taking into account the stochasticity inherent in the transmission dynamics. We can also emphasize the importance of using models that do not imply mass-action mixing of individuals in the population.

Finally, Figures 7 and 8 compare, respectively, the estimates and point forecasts of chickenpox infections obtained with the proposed model, the SMDT-DTR model and the DMCT model. As expected, both the SMDT-DTR and the DMCT models, which consider a seasonal pattern that is constant over time, are not able to properly describe
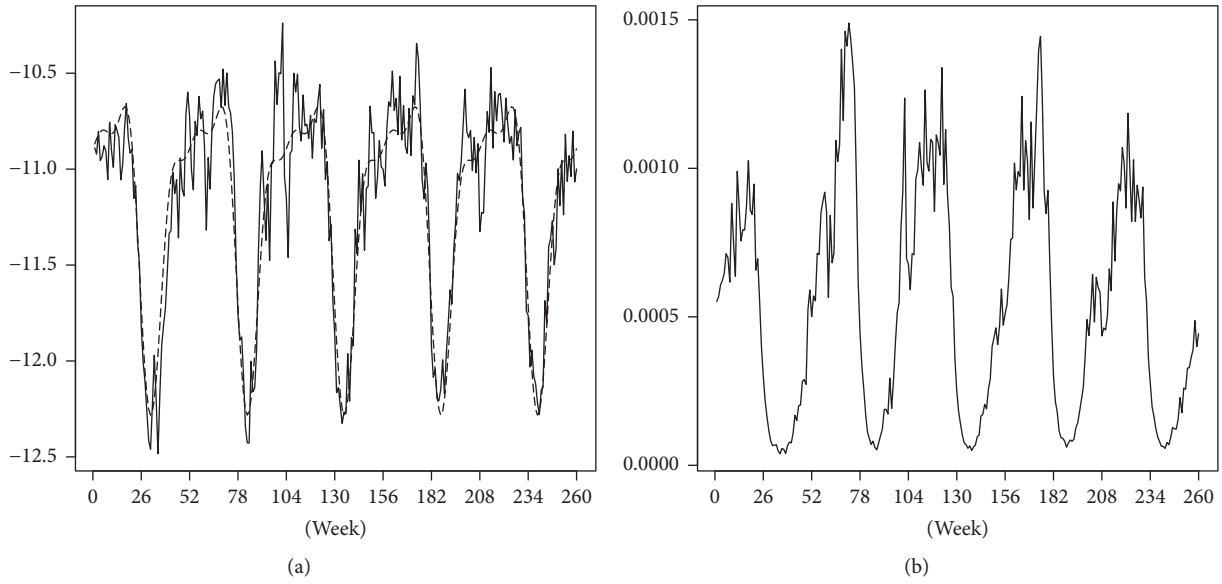
(a)

(b)

Figure 4: Estimated transmission rate together with its seasonal component (a) and estimated probability of getting the virus (b) from week 2, 2007, to week 1, 2012.
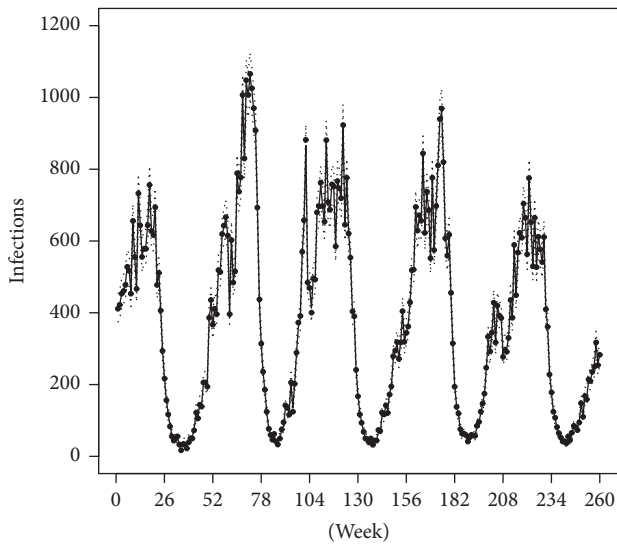

Figure 5: Reported cases of chickenpox (solid points) together with posterior means (solid line) and 95% credible intervals (dotted line) from week 4, 2007, to week 3, 2012.
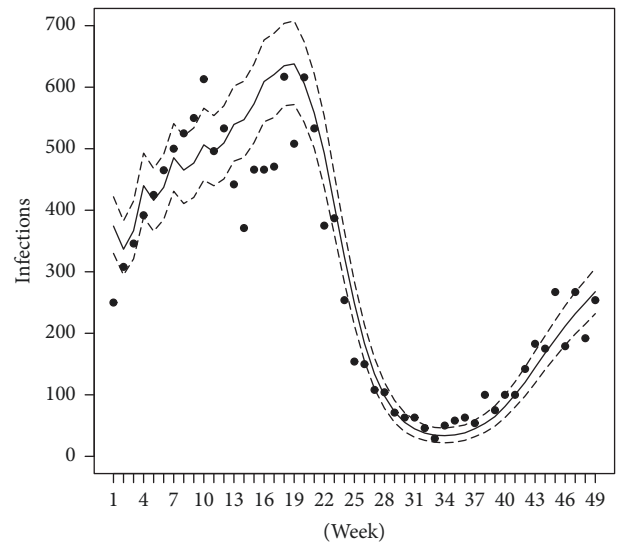

Figure 6: Point forecasts (solid line) together with 95% prediction intervals (dashed lines) for the last 49 weeks of 2012 (from week 4 to week 52, 2012). Solid points represent cases of chickenpox reported by the health sentinel network in the Valencian Community.

annual epidemics. Besides, in this case study, the commonly used deterministic continuous-time approach cannot explain properly epidemic peaks.

5. Conclusions

We have proposed a flexible stochastic compartmental model in discrete time to understand the dynamics of chickenpox, which is an infectious disease with a latency period. Unlike standard formulations, the proposed model does not imply homogeneous mixing of individuals in the population and

takes into account the stochasticity inherent in the transmission of disease, providing a much more realistic and accurate description of the progression of the disease. In addition, the Bayesian analysis of the model allows us to quantify uncertainty about the parameters of the model and facilitates computation of the posterior predictive distribution. The model closely predicts the timing and magnitude of the chickenpox epidemic in 2012.

Although the results obtained are encouraging, we must note that the proposed model has some limitations. We have
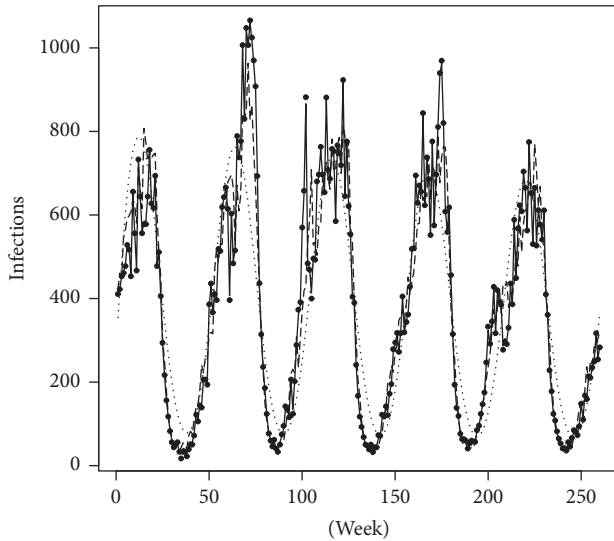
FIGURE 7: Reported cases of chickenpox from week 4, 2007, to week 3, 2012 (solid points), and the estimates obtained with the Bayesian stochastic SEIR model in discrete time proposed in this paper (solid lined), the stochastic model in discrete time with a deterministic seasonal transmission rate (dashed line) and the deterministic continuous-time model with a seasonal transmission rate and heterogeneous mixing (dotted line).
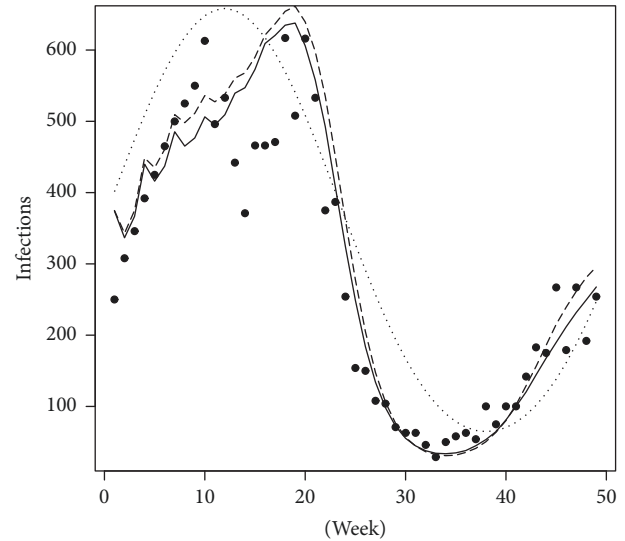

FIGURE 8: Reported cases of chickenpox from week 4 to week 52, 2012 (solid points), and point forecasts obtained with the Bayesian stochastic SEIR model in discrete time proposed in this paper (solid lined), the stochastic model in discrete time with a deterministic seasonal transmission rate (dashed line) and the deterministic continuous-time model with a seasonal transmission rate and heterogeneous mixing (dotted line).

TABLE 2: Fitting and forecast RMSE for the proposed model and four alternative formulations of the SEIR model: a stochastic model in discrete time with a deterministic seasonal transmission rate, a deterministic model in discrete time with a seasonal transmission rate, and two deterministic continuous-time formulations, one with heterogeneous and the other with homogeneous mixing of individuals.

Model	Fitting RMSE	Forecast RMSE
Proposed model	4.41	64.24
SMDT-DTR	89.08	73.32
DMDT	106.79	80.27
DMCT	135.09	106.71
DMCT-Homo	229.26	309.89

assumed that the number of births equals the number of deaths so that the total population size is constant over time. In addition, we do not consider an age structure into the formulation of the compartmental model. It would be valuable to extend the proposed model to allow for different age groups. An age-structured model would also provide an important tool to study the effects of alternative vaccination strategies.

Another very fruitful area for further research is the extension of the model into the spatial domain. Space can play a significant role in disease dynamics. In addition, a spatiotemporal model would help to detect high-risk areas in need of more strong intervention strategies to reduce the burden of disease. The incorporation of covariates affecting disease transmission, such as temperature or wind speed, may also improve description and prediction of the pattern of disease.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] U.S. Department of Health and Human Services, "Centers for Disease Control and Prevention. Chickenpox (Varicella)," http://www.cdc.gov/chickenpox/about/index.html.

[2] F. Brauer, "Compartmental models in epidemiology," in *Mathematical Epidemiology*, F. Brauer, P. van den Driessche, and J. Wu, Eds., Springer, Berlin, 2008.

[3] Z. Ma and J. Li, "Dynamical modeling and analysis of epidemics," *World Scientific*, London, 2009.

[4] L. Acedo, J.-A. Moraño, F.-J. Santonja, and R.-J. Villanueva, "A deterministic model for highly contagious diseases: the case of varicella," *Physica A: Statistical Mechanics and its Applications*, vol. 450, pp. 278–286, 2016.

[5] A. Díez-Gandía, R.-J. Villanueva, J.-A. Moraño, L. Acedo, J. Mollar, and J. Díez-Domingo, "Studying the herd immunity effect of the varicella vaccine in the community of Valencia, Spain," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 9656, pp. 38–46, 2016.

[6] V. Isham, "Celebrating statistics," in *Papers in honour of Sir David Cox on his 80th birthday*, A. C. Davison, Y. Dodge, and N. Wermuth, Eds., Oxford University Press, 2004.

[7] L. J. S. Allen and P. van den Driessche, "Stochastic epidemic models with a backward bifurcation," *Mathematical Biosciences and Engineering*, vol. 3, no. 3, pp. 445–458, 2006.

[8] M. Roberts, V. Andreasen, A. Lloyd, and L. Pellis, "Nine challenges for deterministic epidemic models," *Epidemics*, vol. 10, pp. 49–53, 2015.

[9] A. Corberán-Vallet and F. J. Santonja, "A Bayesian SIRS model for the analysis of respiratory syncytial virus in the region of Valencia, Spain," *Biometrical Journal*, vol. 56, no. 5, pp. 808–818, 2014.

[10] O. N. Bjørnstad, B. F. Finkenstädt, and B. T. Grenfell, "Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model," *Ecological Monographs*, vol. 72, no. 2, pp. 169–184, 2002.

[11] A. Gelman, "Prior distributions for variance parameters in hierarchical models," *Bayesian Analysis*, vol. 1, no. 3, pp. 515–533, 2006.

[12] Red Nacional de Vigilancia Epidemiológica and Centro Nacional de Epidemiología. ISCIII, "Centro Nacional de Epidemiología. ISCIIII. nforme sobre la situación de la Varicela y del Herpes Zóster en España," 1998-2012, Madrid, Spain, 2014, http://www.isciii.es/ISCIII/es/contenidos/fd-servicios-cientifi-co-tecnicos/fd-vigilancias-alertas/fd-enfermedades/fd-enfermedades-prevenibles-vacunacion/InformeVaricela_HZ_1998-2012.pdf.

[13] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best, "Rejoinder to commentaries on 'The BUGS project: evolution, critique and future directions' [MR2750401]," *Statistics in Medicine*, vol. 28, no. 25, pp. 3081-3082, 2009.