ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2018)
67, *Part* 5, *pp.* 1379–1398

Stratified space–time infectious disease modelling, with an application to hand, foot and mouth disease in China

Cici Bauer

Pfizer, Cambridge, USA

and Jon Wakefield

University of Washington, Seattle, USA

[Received December 2015. Final revision March 2018]

Summary. We extend an interesting class of space–time models for infectious disease data proposed by Held and co-workers, to analyse data on hand, foot and mouth disease, collected in the central north region of China over 2009–2011. We provide a careful derivation of the model and extend the model class in two directions. First, we model the disease transmission between age–gender strata, in addition to space and time. Second, we use our model for inference on effective local reproductive numbers. For the hand, foot and mouth data, for each of the six age–gender strata we consider that transmission is greatest between individuals within the same strata, with also relatively high transmission between individuals of the same age group but the opposite gender. The local reproductive numbers show strong seasonality, and between-area differences.

Keywords: Bayesian inference; Infectious disease; Reproductive numbers; Space–time modelling; Transmission modes

1. Introduction

In this paper we analyse surveillance data on hand, foot and mouth disease (HFMD), collected by the Chinese Center for Disease Control and Prevention. The substantive aim of this analysis was to estimate the transmission rates between six age–gender strata, and to estimate local reproductive numbers. HFMD is an acute contagious infection and is usually seen in children; there have been large-scale outbreaks in Asia (Tong and Bible, 2009). Transmission may occur through person-to-person contact, or contact with contaminated faeces, surfaces and objects. The disease is caused by enterovirus pathogens and usually produces mild or moderate symptoms such as fever, oral ulcers or rashes on the hand, foot and mouth. However, HFMD can cause severe illness with neurological problems and even death. Enterovirus-related HFMD, with the first large-scale epidemic outbreak in 2008 in China, has been included as one of the 39 notifiable infectious diseases in the Chinese Center for Disease Control disease surveillance system. Each reported case from the surveillance system includes information on the person's current home address, gender, age and the symptom onset date. More information about the China HFMD surveillance data can be found in Wang *et al.* (2011).

Address for correspondence: Cici Bauer, Early Clinical Development Non-clinical Statistics, Pfizer Worldwide Research and Development, 1 Portland Street, Cambridge, MA 02139, USA.
E-mail: cici_bauer@brown.edu

 0035–9254/18/671379

In this paper, we focus on data from the central north region of China collected between 2009 and 2011; this region is shown in relation to the whole of China in the on-line supplementary materials. The central north region consists of 59 prefectures in five provinces (Tianjin, Hebei, Henan, Shandong and Shanxi) and one direct-controlled municipality (Beijing). The geographic size of the administrative divisions in China, from the smallest to the largest, is township, county, prefecture and province. We perform our analysis at the prefecture level because of the better quality of reporting at this administrative level. The total population in the region is assumed to be constant over 2009–2011 and is estimated to be 318 022 505 from the national bureau of statistics of China. Within the central north region, 418 949, 478 238 and 311 118 HFMD cases were reported in 2009, 2010 and 2011 respectively. We aggregate the number of HFMD cases by week and by prefecture, and partition the population into three age groups: less than 1, [1,6) and 6 years old or older, which roughly correspond to infants, children in kindergarten and children in school (the vast majority of the cases are in children). Together with gender, the age partition gives six age–gender demographic subgroups; Fig. 1 displays the incidence curves. In all subgroups the HFMD season appears to start around March, reaches its peak in May–June and gradually dies down towards the winter; however, there are differences in both the shape and the magnitude of the epidemics across subgroups.

In recent years, modelling infectious disease data in space and time has gained increasing popularity, partially due to the availability of such data. Often these data are obtained from government disease surveillance systems and in many situations are in the form of spatially and temporally aggregated disease counts. To address the inference problem when fitting compartmental infectious disease models, various approaches have been suggested. These compartmental systems model the numbers of individuals moving through the different disease phases; we utilize a susceptible–infectious–recovered (SIR) compartmental model.

We first briefly summarize inferential approaches that are based on continuous time Markov chains, before describing the discrete time Markov chain models that we concentrate on. For small populations, Bayesian auxiliary variable methods, which impute unobserved infection and recovery times, are computationally practical (Gibson and Renshaw, 1998; O'Neill and Roberts, 1999; O'Neill and Becker, 2001). These models are not feasible in medium-to-large populations, and various approximations have been suggested. A diffusion process approximation was suggested by Cauchemez and Ferguson (2008) whereas particle filtering, for likelihood or Bayesian inference, is a more computationally expensive approach that has been used by He *et al*. (2010) and Koepke *et al*. (2016). Iterated filtering has also been proposed; see for example Ionides *et al*. (2015). Methods for modelling infections in continuous time are appealing, since they reflect a plausible generating mechanism (albeit under the often dubious assumption of homogeneous mixing), but they are not currently computationally feasible for the population sizes of this study. Disease-mapping-type models that do not explicitly acknowledge the infectious aspect have been used by Mugglin *et al*. (2002), Knorr-Held and Richardson (2003) and Bauer *et al*. (2016). These models may be useful for prediction and mapping, but we would like to learn about transmission parameters and so we choose to use discrete time models that have been explicitly developed for infectious diseases. Specifically, the framework that we work with is based on a discrete time Markov chain SIR model, as originally proposed by Held *et al*. (2005). The approach was further extended by Paul *et al*. (2008), Paul and Held (2011), Held and Paul (2012), Meyer and Held (2014) and Geilhufe *et al*. (2014). For an excellent review of the statistical aspects of infectious disease modelling, and this class of models in particular, see Höhle (2016). In this paper we develop this 'epidemic–endemic' framework for our context, by extending to allow stratum-specific transmission. Other related approaches include Lekone and Finkenstädt (2006) who considered a susceptible–exposed–infected–recovered

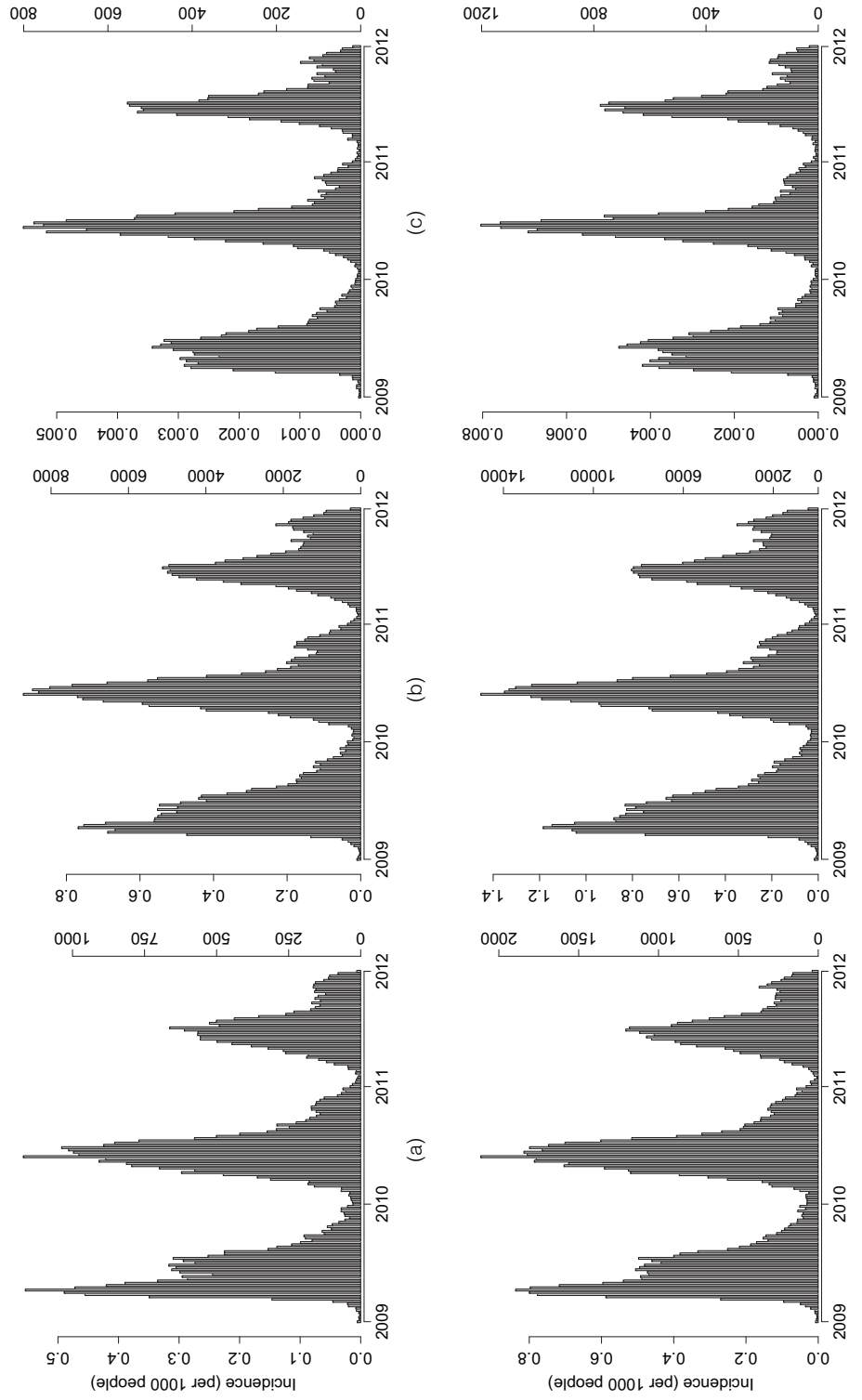

Fig. 1. Weekly epidemic curves of HFMD cases between 2009 and 2011 in the central north region of China, categorized by six age–gender subgroups (in each plot, the left-hand axis shows the rate of incidence and the right-hand axis shows the corresponding absolute number of cases; note the different scaling of the y-axes in each plot): (a) female, < 1 year; (b) female, [1, 6) years; (c) female, $\geqslant$ 6 years; (d) male, < 1 year; (e) male, [1, 6) years; (f) male, $\geqslant$ 6 years

compartmental system and modelled ebola. Time series SIR models are closely related to the epidemic–endemic models and have a long history, beginning with Finkenstädt and Grenfell (2000) and with other statistical descriptions including Bjørnstad *et al*. (2002), Morton and Finkenstädt (2005) and Bjørnstad and Grenfell (2008). For a review of, and comparison between, epidemic–endemic and time series SIR models, see Wakefield *et al*. (2018). Approximate Bayesian computation approaches to the modelling of infectious diseases data have been suggested by McKinley *et al*. (2009) and Toni *et al*. (2010), but the inherent approximations are difficult to assess.

We briefly outline the structure of the paper. In Section 2 we carry out initial analyses of the China HFMD data; this section motivates the need to extend the model to allow for transmission between different strata, as defined by age and gender, to obtain a better understanding of the manner in which the disease transmits through the population. This extension is described in Section 3. Section 4 provides the data analysis of the China HFMD data. We conclude the paper with a discussion in Section 5. The on-line supplementary materials contain more technical details and additional supporting information.

Some simulated data and the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Initial data analysis

In their simplest forms, the Held *et al*. (2005) models can be viewed as Poisson branching processes with immigration. The mean risk is decomposed into an endemic component and an epidemic component, with the latter having two sources of infective individuals: self-area and neighbouring area. These models are implemented within the `surveillance` package in R (Meyer *et al*., 2017). Inference for these models is made by using standard likelihood inference, or penalized quasi-likelihood when random effects are considered; however, inference from the latter can be unreliable in some contexts (Breslow and Clayton, 1993). Notably, the implementation does not provide a straightforward way to allow both age–gender and space to be in the model. The modelling framework that was initiated by Held *et al*. (2005) is implemented in the workhorse function `hhh4` in R (Meyer *et al*., 2017).

We motivate our model extensions with an analysis of the HFMD data using a separate Held *et al*. (2005) model for each of the six age–gender groups. First, let y_{it} be the observed count of HFMD cases in area i and in week t for a generic age–gender group, $i = 1, \ldots, n$ and $t = 1, \ldots, T$. We also let $\mathbf{z}_{it}$ represent a $q \times 1$ vector of area–time-specific bases from which a covariate model may be constructed. These bases may correspond to terms in a spline or change point model or lagged versions of variables, such as temperature. The population in area i is denoted as N_i and is assumed constant over the study period. The estimated incubation period of HFMD is between 3 and 7 days (Chang *et al*., 2012; Huang *et al*., 2013; Wu *et al*., 2014), and so a weekly timescale is chosen (further discussion appears in Section 3). We assume that $Y_{it}|\mu_{it} \sim \text{NegativeBinomial}(\mu_{it}, \psi)$ with conditional mean $\mu_{it} = E[Y_{it}|Y_{i,t-1} = y_{i,t-1}]$ and scale parameter ψ resulting in the variance form $\mu_{it}(1 + \mu_{it}/\psi)$. The mean is decomposed into three terms, which we refer to as auto-regressive AR, neighbourhood NE and endemic EN,

$$\mu_{it} = \lambda_{it}^{\text{AR}} y_{i,t-1} + \lambda_i^{\text{NE}} \sum_{i'=1}^{n} w_{i'i} y_{i',t-1} + \lambda_{it}^{\text{EN}}. \tag{1}$$

The auto-regressive rate λ_{it}^{AR} is a multiplier on the previous week's cases in area i, whereas the neighbourhood rate λ_i^{NE} is a multiplier on the scaled previous week's cases from neighbouring

areas. Various choices are available for the weights and we choose $w_{i'i} = 1/|\text{ne}(i')|$ for $i' \in \text{ne}(i)$, with ne(i) being the set of neighbours of area i. The rationale here is that the cases in neighbouring area i' are distributed to *their* neighbours (which includes area i) in proportion to the number of neighbours. We take areas as neighbours if they share a common geographical boundary. Other neighbourhood weighting schemes are described in Paul *et al.* (2008), Paul and Held (2011), Held and Paul (2012), Meyer and Held (2014) and Geilhufe *et al.* (2014). The so-called endemic component λ_{it}^{EN} in equation (1) is a term that includes all contributions that are not catered for by the auto-regressive and neighbourhood components (including contributions from areas that are not included in the neighbour set); sometimes this component is referred to as the environmental reservoir. In Section 3, we shall motivate the aggregate model (1) via an idealized development from the level of the individual.

The areas that we consider for HFMD are relatively large and so we expect the lion's share of direct transmission to be self-area; consequently, we can afford a more complex model for the associated rate. We specifically assume the log-linear model

$$\log(\lambda_{it}^{\text{AR}}) = \alpha_0^{\text{AR}} + \mathbf{z}_{it}^{\text{T}}\boldsymbol{\beta}^{\text{AR}} + b_i^{\text{AR}}, \tag{2}$$

where $\boldsymbol{\beta}^{\text{AR}}$ is a $q \times 1$ vector of association parameters, and $b_i^{\text{AR}}|\sigma_{\text{AR}}^2 \sim_{\text{IID}} N(0, \sigma_{\text{AR}}^2)$ is an area level random effect. A more complicated form that acknowledges possible spatial dependence can be adopted, e.g. the intrinsic conditional auto-regressive (ICAR) model of Besag *et al.* (1991) or the model of Leroux *et al.* (1999). The neighbouring area rate assumes the simpler form

$$\log(\lambda_i^{\text{NE}}) = \alpha_0^{\text{NE}} + b_i^{\text{NE}}, \tag{3}$$

with neighbourhood random effect $b_i^{\text{NE}}|\sigma_{\text{NE}}^2 \sim_{\text{IID}} N(0, \sigma_{\text{NE}}^2)$. The log-linear model for the endemic component is

$$\log(\lambda_{it}^{\text{EN}}) = \log(N_i) + \alpha_0^{\text{EN}} + b_i^{\text{EN}} + \beta_{\sin}\sin\left(\frac{t}{52}2\pi\right) + \beta_{\cos}\cos\left(\frac{t}{52}2\pi\right), \tag{4}$$

where $b_i^{\text{EN}}|\sigma_{\text{NE}}^2 \sim_{\text{IID}} N(0, \sigma_{\text{EN}}^2)$ is the random effect. The offset accounts for the population, and seasonality is modelled via the sine and cosine terms, as recommended in usual implementations of this model, e.g. Meyer *et al.* (2017).

We now discuss the covariate model for the auto-regressive component. We include an indicator variable for whether school was open, with school closure taken as January 15th–February 15th and July 1st–August 31st, and school open corresponding to the remainder of the year. This definition was taken from Wang *et al.* (2011). The rest of the covariate vector consists of the meteorological variables temperature, precipitation, relative humidity and wind speed. The meteorological data were obtained from the National Climate Data Center of the US Department of Commerce, which includes data from more than 300 weather stations in mainland China. We applied a tensor product cubic regression spline model to the averaged daily weather data from the monitoring stations, to obtain area level summaries such as area averages. The functional form to model the effects of meteorological variables was determined from exploratory analyses; we assumed a change point model for temperature (with the change at 46.6 °F) and linear terms for the remaining variables. Details on this process, and the interpolation procedure, can be found in the on-line supplementary materials, section S3. Given the 3–7-day incubation period of HFMD, it is reasonable to assume that the meteorological conditions should enter the model at a 1-week lag.

The results from fitting model (1), to each of the six age–gender strata, are presented in Table 1. Clearly there is a large variation in the auto-regressive intercept parameters between strata. The covariate associations are more consistently estimated across strata whereas the seasonality

Table 1. Results from fitting the model of Held *et al.* (2005) to each age–gender group, indexed by 1–6, of the China central north weekly HFMD surveillance data†

Parameter	*Results for the following groups:*					
	1, female, < 1 year	*2, female, [1–6) years*	*3, female, ⩾ 6 years*	*4, male, < 1 year*	*5, male, [1, 6) years*	*6, male, ⩾ 6 years*
α_0^{AR}	−2.80	−1.73	−4.22	−2.35	−1.67	−4.26
	(0.22)	(0.081)	(0.39)	(0.16)	(0.069)	(0.34)
α_0^{NE}	−1.60	−2.18	−1.31	−1.78	−2.37	−1.41
	(0.10)	(0.11)	(0.11)	(0.12)	(0.12)	(0.10)
α_0^{EN}	−11.36	−20.39	−16.00	−11.19	−22.08	−12.13
	(0.11)	(0.44)	(0.11)	(0.13)	(0.45)	(0.111)
β_{school}	0.022	0.064	0.18	0.0034	0.070	0.17
	(0.039)	(0.018)	(0.061)	(0.03)	(0.015)	(0.050)
β_{tempr}	0.047	0.031	0.068	0.038	0.031	0.071
	(0.0041)	(0.0014)	(0.007)	(0.0035)	(0.0012)	(0.0067)
β_{tempr2}	−0.058	−0.037	−0.077	−0.0011	−0.036	−0.079
	(0.005)	(0.0019)	(0.0087)	(0.0012)	(0.0016)	(0.0075)
β_{humid}	−0.0032	−0.0009	−0.0041	−0.078	−0.0008	−0.0039
	(0.0013)	(0.00062)	(0.0033)	(0.012)	(0.0058)	(0.0018)
β_{windsp}	0.10	0.049	0.13	0.084	0.053	0.14
	(0.0094)	(0.0067)	(0.023)	(0.0066)	(0.0052)	(0.019)
β_{precip}	0.18	−0.35	0.13	−0.0078	0.057	−0.038
	(0.074)	(0.061)	(0.21)	(0.10)	(0.016)	(0.17)
β_{sin}	0.088	5.64	0.14	0.085	11.98	0.14
	(0.064)	(0.43)	(0.059)	(0.067)	(0.42)	(0.059)
β_{cos}	−0.98	−1.94	−0.91	−1.19	1.21	−0.81
	(0.062)	(0.17)	(0.061)	(0.074)	(0.12)	(0.061)
σ_{AR}	0.29	0.099	0.32	0.26	0.087	0.22
σ_{NE}	0.67	0.75	0.73	0.85	0.78	0.70
σ_{EN}	0.69	0.88	0.80	0.66	0.98	0.75
ψ	6.7	10.0	4.8	8.3	11.2	5.6

†Independent normal random effects are included in each of the three model components, with a set of sine–cosine bases for seasonality. Log-linear terms are assumed for the meteorological variables, with the coefficients for temperature representing the two slopes that meet at the change point of 46.6 °F. The scale parameter of the negative binomial distribution is ψ. The standard errors of the parameter estimates are shown in parentheses.

(as represented by the sine and cosine terms) shows marked variability across strata. It is more difficult to compare variance components, but the random-effect standard deviations do not appear to be wildly different between strata. We also fitted this model with a Poisson distribution for the counts, and the results from this analysis are in the supplementary materials, section S6.1. As expected, the negative binomial standard errors were in general larger than their Poisson counterparts.

This analysis is partially informative on the role of age–gender in transmission; only an analysis in which the counts from all six strata are simultaneously estimated enables the estimation of between-strata transmission rates. For contact-driven infectious diseases, understanding the transmission between various demographic groups is central to understanding the aetiology of the disease and designing effective interventions (e.g. in which groups vaccination strategies concentrate on). Therefore, in our proposed model framework, the ability to consider age–gender strata and space simultaneously is a major extension to previous approaches.

3. Model description

In this section, we describe the Bayesian stratified space–time model that was motivated by the China HFMD data. It was not obvious how to extend the framework of Held *et al*. (2005) to allow for stratification and so we first develop a discrete time chain binomial SIR model (see, for example, Becker (1989)) to motivate a particular form of the model. We then describe the extension to include strata, space and time simultaneously. Additional technical details can be found in the on-line supplementary materials, section S1.

3.1. A discrete susceptible–infectious–removed model

We model the probability that, in a generic area, a susceptible individual at time $t-1$ will become infected by time t. We assume that infected individuals are infectious for 1 time unit, before becoming removed, so that we have an SIR model. Given y_{t-1} infective individuals, the hazard (force of infection) for a susceptible person at time $t-1$ is given by

$$\lambda_t^{\dagger} = c(N)\,\frac{y_{t-1}}{N}\,p,$$

where $c(N)$ is the unit rate of contacts that the susceptible person experiences, y_{t-1}/N is the probability that a randomly selected contact is infectious and p is the per-contact probability of infection. Different infectious disease models treat the time until infection as either continuous or discrete. Clearly time to infection is continuous but often (such as in the HFMD example) the data arise in a form which makes it more convenient to describe as discrete. The development is similar to the discretized susceptible–exposed–infectious–recovered model of Lekone and Finkenstädt (2006); see their equations (5) and (6) for very similar forms to those derived below. Section S1 in the on-line supplementary materials contains a detailed discussion and development of models under various modes of transmission. We discuss the *frequency-dependent* model which is the transmission mode that motivates the model that we fit to the HFMD data. Under this model the contact rate is independent of population size, $c(N) = c^{\mathrm{FD}}$, and the hazard is $\lambda_t^{\dagger} = c^{\mathrm{FD}} p\, y_{t-1}/N = \lambda\, y_{t-1}/N$. The hazard is assumed constant over the 1-unit time interval $(t-1, t]$ so that each susceptible has an exponential time until infection, giving the probability of infection as $1-\exp(-\lambda y_{t-1}/N)$. By discretizing time and assuming a constant hazard in each time interval, we are ignoring both infective individuals becoming recovered, and other susceptible individuals becoming infected in $(t-1, t]$. If we assume that all susceptible individuals at time $t-1$, denoted S_{t-1}, have independent infection times, we have a Reed–Frost chain binomial model (e.g. Daley and Gani (1999), chapter 4) with $Y_t | Y_{t-1} = y_{t-1} \sim \text{binomial}\{S_{t-1}, 1-\exp(-\lambda y_{t-1}/N)\}$. In a large population, and if the disease is rare, so that the susceptible pool is not significantly diminished, we may approximate the number of susceptible individuals by the total population size N. In the case of a rare disease, $\lambda y_{t-1}/N$ is small, and $\exp(-\lambda y_{t-1}/N) \approx 1-\lambda y_{t-1}/N$ so, with susceptible individuals approximated by N and the rare disease assumption, $Y_t | Y_{t-1} = y_{t-1} \sim \text{Poisson}(\lambda y_{t-1})$.

The basic reproductive number, which is usually denoted by R_0, is of central importance in infectious disease epidemiology. It is defined as the average number of secondary cases infected by a typical infected person in a completely susceptible population. The basic reproductive number under the above model is $R_0 = \lambda$. We observe that in this very simple model, if $y_t = 0$ for any t, the chain of infections ends.

3.2. Extension to multiple areas and strata

We stratify the population into age–gender demographic strata and consider multiple areas. Let Y_{itj} be the number of new infections in area i, week t and stratum j, $i = 1, \ldots, n$, $t = 1, \ldots, T$,

$j = 1, \ldots, J$. Infection for a susceptible person in area i and stratum j can occur from an infective person in area i and stratum j', or an infective person in neighbouring area i' and stratum j' (in both cases including $j' = j$), or from the environmental reservoir (i.e. the endemic component).

In the case of competing ways of becoming infected, we can use the classic competing risks framework of Prentice *et al.* (1978), in which the hazards are additive. We let $\lambda_{itj}^{\mathrm{TOT}}$ represent the overall hazard for a stratum j susceptible person in area i at time t, and we write

$$\lambda_{itj}^{\dagger,\mathrm{TOT}} = \lambda_{itj}^{\dagger,\mathrm{AR}} + \lambda_{itj}^{\dagger,\mathrm{NE}} + \lambda_{itj}^{\dagger,\mathrm{EN}},$$

so that the hazard sums over contributions from both the area of the susceptible individual, and neighbouring areas, and from the environment. Following the earlier development, and assuming that $\lambda_{itj}^{\mathrm{TOT}}$ is small, the probability of infection in $[t-1, t)$, for a single susceptible person is

$$1 - \exp(-\lambda_{itj}^{\mathrm{TOT}}) \approx 1 - (1 - \lambda_{itj}^{\mathrm{TOT}}) = \lambda_{itj}^{\mathrm{TOT}}.$$

In the on-line supplementary materials (section S1), we discuss various possibilities for the contact rates that determine the form of the hazard. Here we adopt a form that leads to a natural analogue of the non-stratified model. As we shall see in Section 4, the contribution of the neighbouring component to the fitted mean is small, and so a simple model is suggested; in particular, we assume that the neighbourhood hazard rate is constant over time and identical for all strata. Under these choices, we obtain the mean of Y_{itj},

$$\mu_{itj} = \sum_{j'=1}^{J} \lambda_{itjj'}^{\mathrm{AR}} y_{i,t-1,j'} + \lambda_i^{\mathrm{NE}} \sum_{i'=1}^{n} w_{i'i} \sum_{j'=1}^{J} y_{i',t-1,j'} + \lambda_{itj}^{\mathrm{EN}}, \tag{5}$$

where

(a) $\lambda_{itjj'}^{\mathrm{AR}}$ is the hazard for infection from a single infective person in area i, stratum j' and time t,
(b) $w_{i'i}\lambda_i^{\mathrm{NE}}$ is the hazard of infection from a single infective person in neighbouring area i' and
(c) $\lambda_{itj}^{\mathrm{EN}}$ is the hazard for infection from the environment.

Model (5) can be further extended to allow covariates, area-specific deviations from the overall intercept and/or seasonality by assuming suitable forms for the rates, as we now describe. For the auto-regressive component, we have

$$\log(\lambda_{itjj'}^{\mathrm{AR}}) = \alpha_{jj'}^{\mathrm{AR}} + \mathbf{z}_{it}^{\mathrm{T}} \boldsymbol{\beta}^{\mathrm{AR}} + b_i^{\mathrm{AR}}, \tag{6}$$

where $\alpha_{jj'}^{\mathrm{AR}}$ reflects the transmission between a susceptible person in stratum j and an infective person in stratum j' and the remaining terms are as defined in equation (2). For the less important neighbourhood component, an identical model to that previously specified in equation (3), i.e. $\log(\lambda_i^{\mathrm{NE}}) = \alpha_0^{\mathrm{NE}} + b_i^{\mathrm{NE}}$, is adopted. For the endemic component, we extend equation (4) to include stratum-specific intercepts

$$\log(\lambda_{itj}^{\mathrm{EN}}) = \log(N_{ij}) + \alpha_{0j}^{\mathrm{EN}} + b_i^{\mathrm{EN}} + \beta_{\sin,j} \sin\left(\frac{t}{52} 2\pi\right) + \beta_{\cos,j} \cos\left(\frac{t}{52} 2\pi\right). \tag{7}$$

We note that we have assumed frequency-dependent transmission and that the number of susceptible individuals can be approximated by the population counts N_{ij}, which means that these counts cancel in the auto-regressive and neighbourhood contributions to the mean function (see the supplementary materials, section S1.1, for further detail), but not in the endemic component. To allow for overdispersion, the Poisson model was extended to the negative binomial model

by adding a multiplicative gamma random effect; the scale parameter of the negative binomial is allowed to vary by strata and is denoted by ψ_j, $j = 1, \ldots, J$. Details of the extension can be found in the supplementary materials, section S1.1.

3.3. Estimating local reproductive numbers

Wang *et al.* (2011) estimated local effective reproductive numbers, in an *ad hoc* fashion. We briefly summarize a formal derivation here, which produces a different form, with the on-line supplementary materials, section S1.2, containing more detail. It is more troublesome to define the local (i.e. area-specific) reproductive numbers for the model that contains a neighbourhood component but, as we have noted, the neighbourhood contribution for the HFMD data is small, and so we shall report local reproductive numbers from a model that contains only terms that are auto-regressive and endemic. The latter do not contribute to the reproductive numbers, since they are presumed to be infections arising from an environmental reservoir. Let $\mathbf{y}_{it}$ be the $J \times 1$ vector of counts that has jth element y_{itj}, $j = 1, \ldots, J$. To represent the non-environmental contribution to risk, we write

$$\mathbf{A}_{it}\mathbf{y}_{i,t-1},$$

where $\mathbf{A}_{it}$ is the $J \times J$ matrix containing elements (j, j') given by $\lambda^{\mathrm{AR}}_{itjj'}$, $j, j' = 1, \ldots, J$. Following Diekmann *et al.* (2012) the local reproductive number is given by the largest eigenvalue of $\mathbf{A}_{it}$. This is the long-term average per time period multiplication number. More formally, $R_{0it} = \lim_{m\to\infty} ||\mathbf{A}^m_{it}||^{1/m}$, the spectral radius of $\mathbf{A}_{it}$, which corresponds to the largest eigenvalue of $\mathbf{A}_{it}$. Since $\mathbf{A}_{it}$ is simply a function of the unknown parameters, we can easily make inference about R_{0it} by transforming the samples in the simulation-based approach to inference that we adopt. For the HFMD data we refer to local reproductive numbers as 'effective' since intervention measures may have been introduced to distort the reproductive numbers that would be seen in a completely susceptible population.

3.4. Inference

We take a Bayesian approach to inference and so require prior distributions for the unknown parameters. The log-transmission-rates, $\alpha^{\mathrm{AR}}_{jj'}$ and log-NE and EN intercepts, i.e. α^{NE}_0 and $\alpha^{\mathrm{EN}}_{0j}$, are all well estimated and so we use normal priors with large variances. In situations in which the data are less informative, one might simulate data under proposed priors, to examine whether the range of observables appears reasonable. Simulating from the prior on $\alpha^{\mathrm{AR}}_{jj'}$, $j, j' = 1, \ldots, J$, and finding the largest eigenvalue of the $J \times J$ matrix $\mathbf{A}$, with elements $\exp(\alpha^{\mathrm{AR}}_{jj'})$, gives the induced prior on the reproductive number, which will also be useful. For the elements of the association parameters, $\boldsymbol{\beta}^{\mathrm{AR}}$, we adopt independent normal priors with large variances. For a generic random-effects standard deviation, we assume that $\sigma \sim \mathrm{Unif}(0, U)$ with U a predefined large number (which we took as 10); see Gelman and Hill (2006) for further discussion.

The prior choice on the scale parameter ψ of the gamma distribution requires careful thought. We follow the procedure that is described in Wakefield (2013), chapter 3, and use a log-normal prior for ψ, leveraging the negative binomial mean–variance relationship. Details of the prior specification for ψ, including R functions for prior calculation, can be found in the on-line supplementary materials, section S5, but, briefly, a range for the variance at certain mean values is specified and this in turn leads to a range for ψ.

Whereas the auto-regressive and neighbourhood rates are directly comparable, the endemic term is of very different form, and so it is difficult to assess the relative contributions directly

from the parameters alone. Following Meyer *et al.* (2017) we plot, for specific areas, fitted curves, along with the real data, showing the different absolute contributions. In addition, to quantify the contribution from the three components over time, we define μ_t^{AR}, μ_t^{NE} and μ_t^{EN} as

$$\mu_t^{\mathrm{AR}} = \sum_{i=1}^{n} \sum_{j=1}^{J} \left(\sum_{j'=1}^{J} \lambda_{itjj'}^{\mathrm{AR}} y_{i,t-1,j'} \right), \tag{8}$$

$$\mu_t^{\mathrm{NE}} = \sum_{i=1}^{n} \sum_{j=1}^{J} \left(\sum_{i'=1}^{n} w_{i'i} \sum_{j'=1}^{J} \lambda_i^{\mathrm{NE}} y_{i',t-1,j'} \right), \tag{9}$$

$$\mu_t^{\mathrm{EN}} = \sum_{i=1}^{n} \sum_{j=1}^{J} \lambda_{itj}^{\mathrm{EN}}. \tag{10}$$

The relative proportions from the respective components are then $\mu_t^{\mathrm{AR}}/(\mu_t^{\mathrm{AR}} + \mu_t^{\mathrm{NE}} + \mu_t^{\mathrm{EN}})$, $\mu_t^{\mathrm{NE}}/(\mu_t^{\mathrm{AR}} + \mu_t^{\mathrm{NE}} + \mu_t^{\mathrm{EN}})$ and $\mu_t^{\mathrm{EN}}/(\mu_t^{\mathrm{AR}} + \mu_t^{\mathrm{NE}} + \mu_t^{\mathrm{EN}})$.

Our model does not fit into the framework that is required of the integrated nested Laplace approximation (Rue *et al.*, 2009), and so we resort to Markov chain Monte Carlo (MCMC) methods and use WinBUGS (Spiegelhalter *et al.*, 1998). The advantage of a simulation-based Bayesian framework is that any quantity that is a function of the model parameters, such as the local effective reproductive numbers and the relative contributions, can be easily inferred. This is a significant advantage over estimation procedures that base inference on asymptotic arguments.

For model comparison, we use the deviance information criterion (DIC) (Spiegelhalter *et al.*, 2002) based on the posterior distribution of the deviance statistic. Models with smaller values of DIC are considered a better fit. We try not to overinterpret in particular small differences between models, since it has been shown to underpenalize complex models (Plummer, 2008). Paul and Held (2011) described how one-step predictions and proper scoring rules may be used for model comparison. Unfortunately, if used with MCMC sampling, this is computationally intensive, although more efficient approaches to calculation have been proposed (Held *et al.*, 2010). We assess local fit by using deviance and Pearson residuals.

4. Application to China hand, foot and mouth disease data

4.1. The model

Model (5) represents a general form, and we fit four variants to the weekly HFMD data in 59 prefectures of the central north region of China. We assume a negative binomial sampling model, with six age–gender bands to give $J = 6$ strata. Together with the number of prefectures and weeks, there are a total of $nTJ = 59 \times 156 \times 6 = 55224$ observations in the analysis. As discussed in Section 1, questions of interest include quantifying the transmission between strata and producing area- and time-specific local effective reproductive numbers.

We concentrate on four variants of equation (5).

(a) Model 1 has random effects, $b_i^{\mathrm{EN}} | \sigma_{\mathrm{EN}}^2 \sim_{\mathrm{IID}} N(0, \sigma_{\mathrm{EN}}^2)$ in the endemic component only.
(b) Model 2 adds auto-regressive random effects, $b_i^{\mathrm{AR}} | \sigma_{\mathrm{AR}}^2 \sim_{\mathrm{IID}} N(0, \sigma_{\mathrm{AR}}^2)$, to model 1.
(c) Model 3 adds neighbourhood random effects, $b_i^{\mathrm{NE}} | \sigma_{\mathrm{NE}}^2 \sim_{\mathrm{IID}} N(0, \sigma_{\mathrm{NE}}^2)$, to model 2.
(d) Model 4 adds endemic ICAR spatial random effects, $s_i^{\mathrm{EN}} | \sigma_{\mathrm{EN}}^{\mathrm{ICAR}} \sim \mathrm{ICAR}(\sigma_{\mathrm{EN}}^{\mathrm{ICAR}})$, to model 3.

For each model, we ran a total of 150000 MCMC iterations with the first 10000 discarded

Table 2. Results from fitting the proposed models with China central north weekly HFMD surveillance data using six age–gender groups†

Parameter	*Results for model 1: fixed AR, fixed NE, IID EN*	*Results for model 2: IID AR, fixed NE, IID EN*	*Results for model 3: IID AR, IID NE, IID EN*	*Results for model 4: IID AR, IID NE, ICAR + IID EN*
α^{NE}	−6.41 (−6.51, −6.32)	−6.38 (−6.48, −6.28)	−5.59 (−5.94, −5.35)	−6.12 (−6.39, −5.84)
$\alpha^{\mathrm{EN}}_{\mathrm{F},<1}$	−18.33 (−20.29, −16.18)	−17.67 (−19.58, −15.79)	−19.14 (−21.23, −16.93)	−18.77 (−20.25, −16.59)
$\alpha^{\mathrm{EN}}_{\mathrm{F},[1,6)}$	−16.34 (−17, −15.64)	−16.26 (−16.99, −15.68)	−16.86 (−17.92, −16.2)	−16.88 (−17.64, −16.25)
$\alpha^{\mathrm{EN}}_{\mathrm{F},\geqslant 6}$	−21.4 (−24.37, −19.77)	−21.6 (−24.68, −19.92)	−22.2 (−27.76, −20.04)	−21.85 (−24.24, −19.75)
$\alpha^{\mathrm{EN}}_{\mathrm{M},<1}$	−11.34 (−11.59, −11.1)	−11.33 (−11.56, −11.13)	−11.45 (−11.7, −11.23)	−11.35 (−11.51, −11.19)
$\alpha^{\mathrm{EN}}_{\mathrm{M},[1,6)}$	−14.46 (−15, −13.79)	−14.25 (−14.94, −13.63)	−15.18 (−15.8, −14.71)	−14.94 (−15.44, −14.54)
$\alpha^{\mathrm{EN}}_{\mathrm{M},\geqslant 6}$	−20.7 (−22.42, −19.23)	−20.99 (−24.23, −19.39)	−22.24 (−25.08, −19.94)	−22.18 (−24.99, −19.82)
β_{school}	0.044 (0.029, 0.059)	0.047 (0.033, 0.063)	0.049 (0.031, 0.065)	0.054 (0.043, 0.068)
β_{tempr}	0.026 (0.025, 0.028)	0.027 (0.027, 0.027)	0.026 (0.025, 0.0261)	0.0247 (0.024, 0.0248)
β_{tempr2}	−0.032 (−0.033, −0.03)	−0.032 (−0.033, −0.032)	−0.03 (−0.032, −0.0298)	−0.029 (−0.0291, −0.0289)
β_{humid}	−0.0017 (−0.0016, −0.0013)	−0.0012 (−0.0014, −0.0009)	−0.0014 (−0.002, −0.0009)	−0.0026 (−0.0028, −0.0024)
β_{windsp}	0.046 (0.042, 0.051)	0.049 (0.045, 0.052)	0.046 (0.041, 0.052)	0.041 (0.039, 0.045)
β_{precip}	0.006 (−0.042, 0.056)	−0.014 (−0.066, 0.032)	0.002 (−0.044, 0.056)	0.05 (0.01, 0.095)
σ_{AR}	—	0.027 (0.019, 0.036)	0.03 (0.018, 0.039)	0.021 (0.01, 0.03)
σ_{NE}	—	—	0.98 (0.80, 1.22)	1.25 (0.99, 1.62)
σ_{EN}	0.79 (0.65, 0.97)	0.79 (0.66, 0.98)	0.90 (0.75, 1.10)	0.57 (0.13, 0.78)
$\sigma^{\mathrm{ICAR}}_{\mathrm{EN}}$	—	—	—	0.52 (0.86, 1.60)
DIC	234712	234640	*234072*‡	234343

†Female and male are labelled as F and M respectively. The table shows the posterior median and the 95% posterior credible interval (in parentheses), and the DIC value. In all models, an independent random-effect term is included in the endemic component. Model 4 includes an additional spatial random effect in the endemic component. The between-strata transmission parameters $\exp(\alpha_{jj'})$ are displayed in Fig. 4 in Section 4.2.
‡Best-performing model according to the smallest DIC value.

as burn-in, and we retained every 10th sample in the remaining samples for inference. The convergence of the Markov chains was checked through visual inspection, with selected chains presented in the on-line supplementary materials, section S9. The run times were about 6 h per model, on an Intel Core i3 laptop.

The key parameters and DIC for models 1–4 are displayed in Table 2. Results of the complete parameters and effective number of parameters can be found in the supplementary materials, section S6.8. We parameterized the negative binomial model as a Poisson sampling model with gamma random effects, which increases the number of random effects in the model, and should

be borne in mind when effective numbers of parameters are interpreted. The DIC clearly favours model 3, with the DIC being far smaller than for competitor models. Hence, unless otherwise stated, reported summaries are from model 3.

Fig. 2 examines, for four representative areas, the fitted *versus* the observed counts (summed over strata), colour coded by the contribution from the auto-regressive, neighbourhood and endemic components. Visually we note that, overall, the fit is satisfactory. The contribution from each of the three components varies by area and over time. The estimated proportions to the mean as contributed by the three components, based on equations (8)–(10), are shown in Fig. 3. Over 80% of the risk of infection is contributed by the auto-regressive component during the non-winter seasons, whereas the endemic component contributes the most during the winter season.

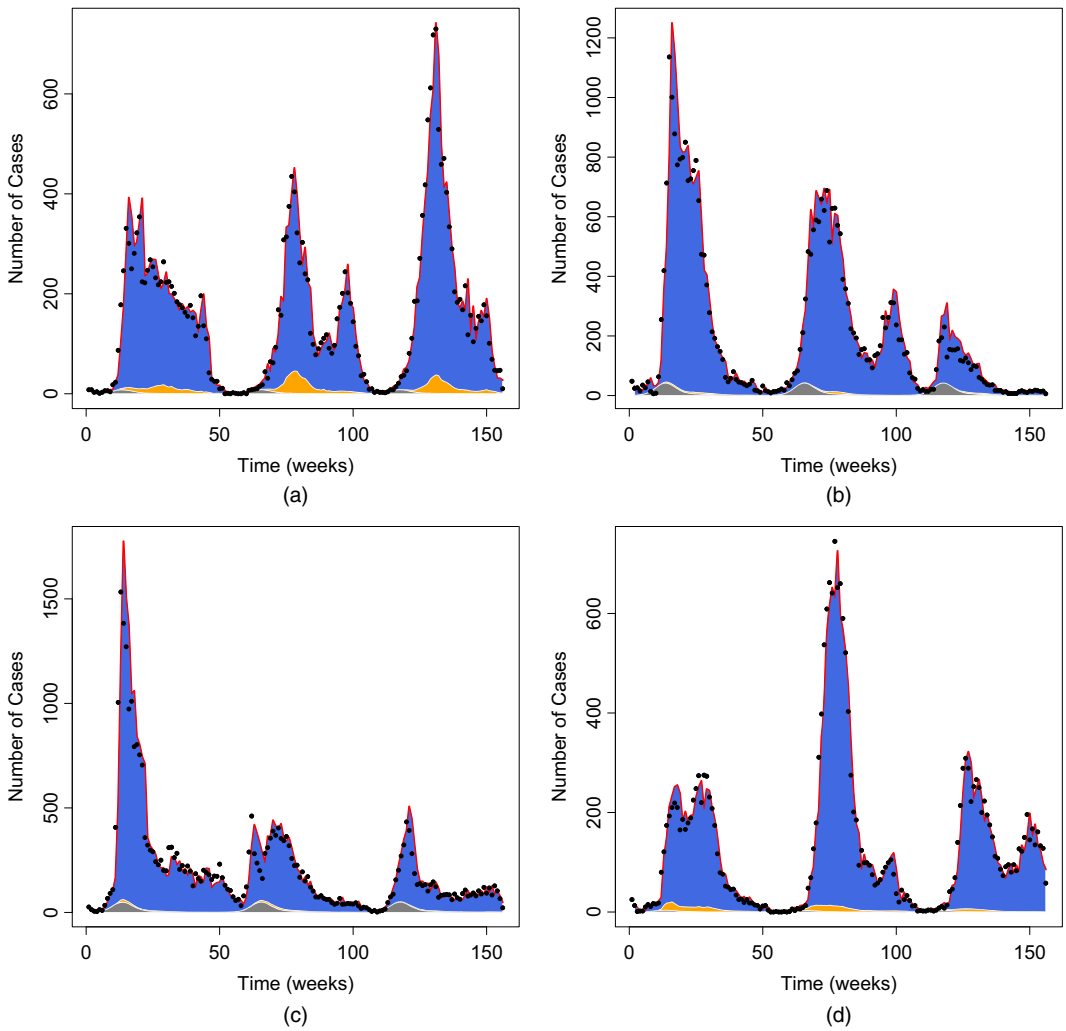

Fig. 2. Fitted (——) and observed (•) numbers of cases, with contributions by the auto-regressive (——), neighbourhood (——) and endemic (——) components, for four provinces in the China central north region (the counts are summed over strata): (a) area 20; (b) area 30; (c) area 47; (d) area 59

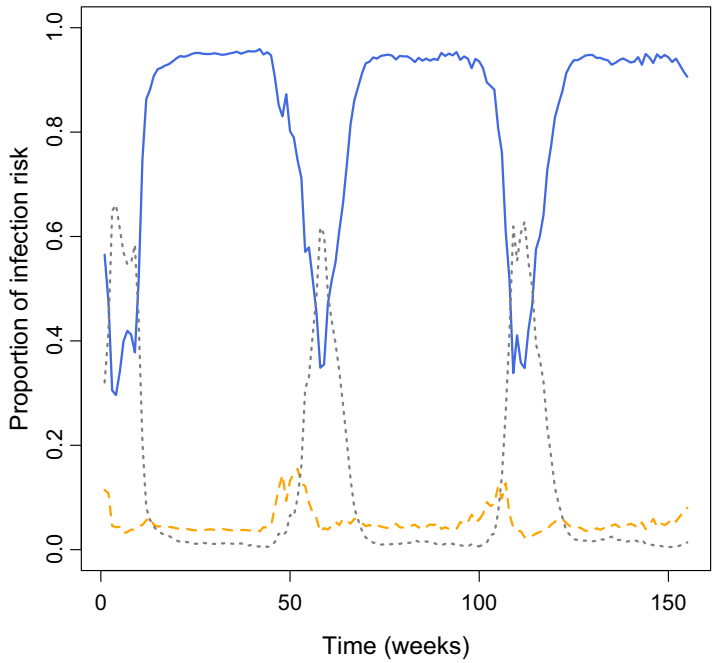

Fig. 3. Estimated proportion of infection risk contributed from the auto-regressive (———), neighbourhood (– – –) and endemic (- - - -) components, from model 3

4.2. Fixed effects

Parameter estimates for the fixed effects from models 1–4 are summarized in Table 2, in which we report the posterior median with 95% posterior credible intervals (CIs) in parentheses. For the intercept parameters $\alpha_{jj'}^{\mathrm{AR}}$, α^{NE} and α_j^{EN}, $j = 1, \ldots, 6$, the estimates are very similar across the four models.

Fig. 4 gives a heat map of the parameters $\exp(\alpha_{jj'})$, which describes the rate of self-area transmission to a stratum j susceptible person from a stratum j' infective person, in a typical area (i.e. one in which the AR random effects are 0). The numerical values appear in Table 8 of the on-line supplementary materials, section S6.6. The results reveal some interesting findings. First, the transmission rates between a susceptible and an infective person show significant within-group, as well as between-group, variation. The banded structure in Fig. 4 is apparent; for all strata, a susceptible person will be infected by an infective person in their same age–gender stratum with the highest transmission rate. For both the boys and the girls of ages less than 1 year, the transmission is dominated by same age strata infective individuals, with same gender infective individuals having slightly higher rates than opposite gender infective individuals. For the oldest groups, the pattern is the same. For the middle age groups, there is more heterogeneity. For example, for boys in the age range $[1, 6)$ years the transmission rates are 0.20 (95% CI [0.19, 0.21]) from same-age boys, 0.16 (95% CI [0.14, 0.17]) from same-age girls and 0.13 (95% CI [0.12, 0.14]) from boys age less than 1 year.

Considering the rates $\exp(\alpha_{jj'})$ as the elements of a $J \times J$ table, we can condition on rows or columns to give two further informative summaries; these are reported in the supplementary materials, section S6.6. Conditioning on rows gives the conditional probabilities that, if a susceptible stratum j individual is infected, the infection arises from a stratum j' infective individual. For example, for a female of age less than 1 year, if a susceptible girl is infected, the

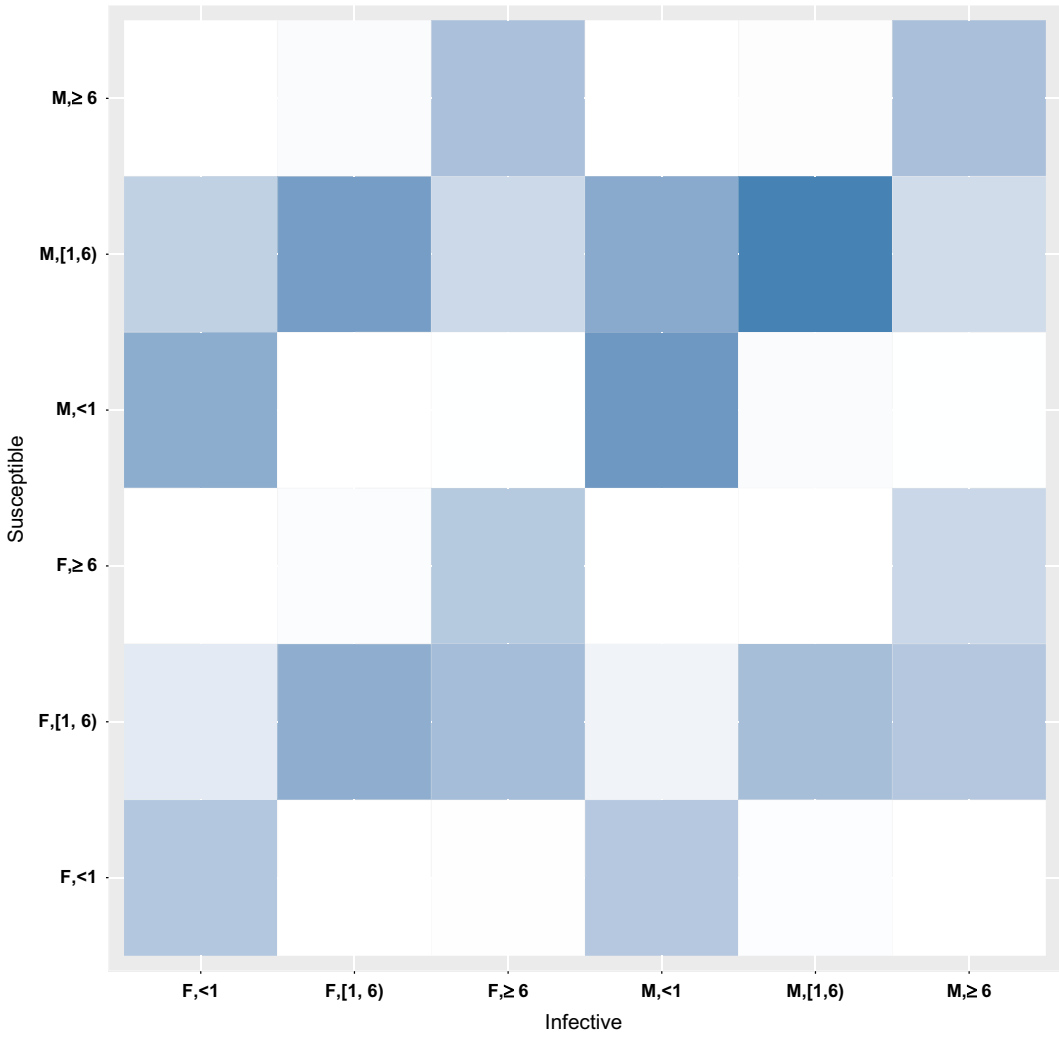

Fig. 4. Estimated transmission rates $\exp(\alpha_{jj'})$ between age–gender strata from model 3 by using China central north weekly HFMD surveillance data: , 0.05; , 0.10; , 0.15; , 0.20

probability that the infection is from a female of the same age group is 0.49, and from a male of the same same age group is 0.43.

The ability to quantify such subgroup transmission mechanisms is important, as it suggests that intervention programmes that are tailored to particular subgroups may be the most effective in controlling and ultimately stopping the spread of the diseases. Yamin *et al.* (2016) estimated transmission rates for different age groups, in the context of respiratory syncytial virus, and discussed how this could inform vaccination strategies.

In all models, 'school in session' is found to be positively associated with infection. For example, in model 3, $\hat{\beta}_{\text{school}} = 0.049$, with 95% CI (0.031, 0.065). This variable can be considered as a proxy for an increased contact rate of about 5% when schools are open. Temperature has a significant effect on the transmission in all the models: the transmission risk increases with temperature until 47 °F (which was the change point in the change point regression model for

temperature), and decreases afterwards. This is an interesting finding and may suggest that the viruses that cause HFMD have an optimal survival temperature. A non-monotonic relationship between temperature and risk of HFMD was also found by Wu *et al*. (2014); see their Fig. 4. Relative humidity is found to have a negative effect on the disease transmission rate, though the effect was very small. Wind speed was found to have a positive effect on the transmission; a higher wind speed is likely to promote the spread of virus in the air and hence increases the risk of contracting the disease. Precipitation had a small and non-significant effect. In Wang *et al*. (2011), a similar relationship with temperature was found, with a higher risk of transmission being associated with temperatures in the range 70–80 °F. The endemic seasonality shows different patterns between the strata, as seen in the estimated coefficients in Table 2 and in the figure contained in the on-line supplementary materials, section S3. The amplitude in particular shows great variability. This is further evidence of the need to include strata information in the model.

4.3. Random effects

The random effects that are associated with the auto-regressive and neighbourhood components are comparable since they both appear as area-specific multipliers on a count; in contrast the endemic random effects appear within a log-linear term without a count multiplier. The variation of the random effects in the auto-regressive component is small compared with the random effects in the neighbourhood component, with the posterior median for σ_{AR} being in the range 0.02–0.03 in models 2–4. The between-area variability of transmission resulting from the neighbouring areas was larger, with σ_{NE} estimated as 0.98 and 1.25 in models 3 and 4 respectively. The intercept for the neighbourhood model is very small, however, and so the neighbourhood component is in general a small contributor to the fit. The variance of the independent endemic random effects was estimated to be 0.90^2 in model 3, suggesting that there is still substantial large residual variability that is unaccounted for in the endemic component. Additional results, including maps of the estimated random effects from the various model components, as well as residual plots, are relegated to the on-line supplementary materials, section S6.4.

4.4. Effective local reproductive numbers

We calculated effective local reproductive numbers, from the no-neighbourhood version of model 3. The parameter estimates are very similar between the model with and without the neighbourhood component; see the on-line supplementary materials, section S6.3, for more details. The reproductive numbers are shown in Fig. 5 for selected areas. The temporal pattern of the reproductive numbers reflects the pattern in the observed counts in Fig. 1, and the temporal variability is largely due to the time varying covariates in the model. Aggregated over all areas, the average reproductive number was around 0.9 and ranged from 0.3 to 1.4 across areas. The estimates are slightly lower than those in Wang *et al*. (2011), where the reproductive numbers were estimated to range between 1.4 and 1.6 (with median 1.4) in spring and to stay below 1.2 in other seasons, among all prefectures in China. We also noted a substantial area-to-area variation in reproductive numbers. Maps of the local reproductive numbers can be found in the supplementary materials, section S6.7. In an on-going surveillance situation, monitoring local reproductive numbers enables public health authorities to decide on areas in which vaccination campaigns should be concentrated. An obvious strategy is to take action when the reproductive numbers pass a threshold, though total numbers of infective cases are also an important input to decision making. A prospective scheme would update parameter estimates on some schedule (weekly, for example), to attempt to highlight in a more timely fashion areas that are going into epidemic phases.

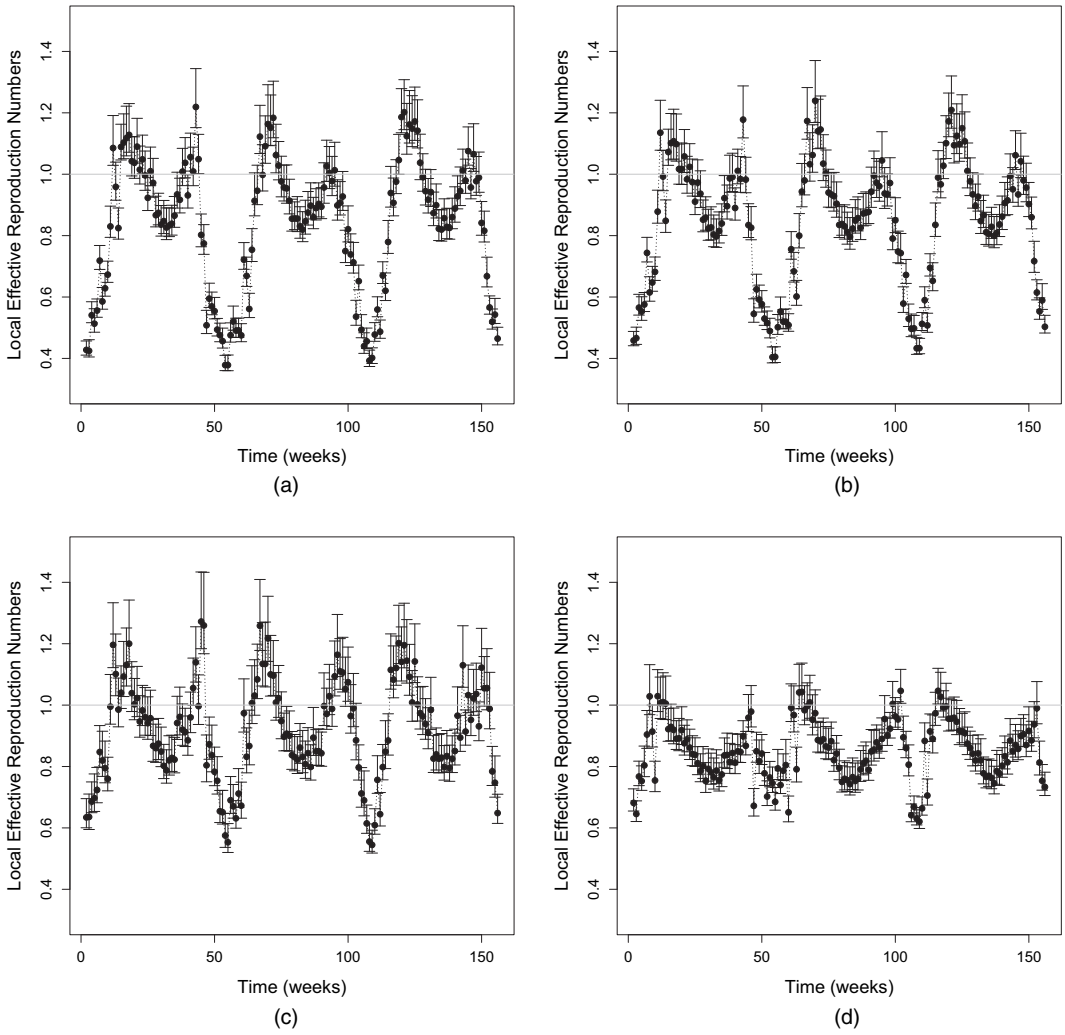

Fig. 5. Estimated local effective reproductive number R_{0it} over time t, for China HFMD surveillance data, for selected areas i from model 3 (▒, posterior 95% CIs; the peaks of the reproductive number reflect the peaks of the observed counts): (a) area 6; (b) area 23; (c) area 26; (d) area 36

5. Discussion

In this paper we have described an extension of the model of Held *et al*. (2005) that enables inference to be made on infectious disease dynamics with respect to time, space and strata. Such a model is potentially useful for many infectious disease applications, given the increasing availability of aggregated space–time data with subgroup information. Routinely collected surveillance data provide one such example and, here, we carried out an analysis of HFMD disease in China.

Bayesian model fitting was carried out for inference. One advantage of this approach is that the analysis is essentially just as straightforward in the presence of missing observations; for example, counts may not be observed in all time periods or there may be short-term failures in the reporting system which may produce sequences of missing values. This is a critical problem

in auto-regressive-type models as one missing time period has a knock-on effect in subsequent periods. With an MCMC implementation one simply includes the missing counts as auxiliary variables. With the frequentist approach that was adopted in Held *et al.* (2005) (and subsequent references) it is not straightforward to deal with such missing data. Informative summaries, such as local reproductive numbers, can be computed easily from the joint posterior distribution. We utilized the well-known WinBUGS software but the computation is demanding. We have also programmed simple versions of the models that are presented here in Stan (Carpenter *et al.*, 2016) and this development has been continued for more complex models; Wakefield *et al.* (2018).

Another contribution of this paper is a derivation of the model form from an individual level model (based on a simple Reed–Frost chain binomial form). This allows the final form assumed to be critically assessed in terms of the assumptions of this development; for example, different competing forms correspond to differing assumptions about the modes of transmission.

For the China HFMD data, we assumed a 1-week incubation period for the infection, and assumed that a susceptible person in week t was potentially infected by infective individuals from week $t-1$. We describe the extension to the more general case for a generic area and stratum, and consider the auto-regressive component only. To allow for infection from infective individuals from up to k ($\geqslant 1$) weeks previously we have

$$E[Y_t|Y_{t-1}=y_{t-1},Y_{t-2}=y_{t-2},\ldots,Y_{t-k}=y_{t-k}]=\sum_{l=1}^{k}\lambda_l^{\mathrm{AR}}q_l y_{t-l}, \tag{11}$$

where the λ_l^{AR}s are the hazard rates that are associated with infective individuals from week l, and q_l is the probability that a case from l weeks before is infective. In general, how many previous weeks to consider and the associated probability q_l are difficult to estimate from the data (Wang *et al.*, 2011).

For enterovirus 71 HFMD the basic reproduction number has been estimated as 5.5 and for Coxsackievirus A16 HFMD as 2.5 (Ma *et al.*, 2011). We do not have cases broken down by pathogen type, but the numbers that we estimated are smaller than these values. One possible source of the disparity is that the method of Ma *et al.* (2011) was based on the cumulative number of cases in the initial growth phase of the outbreaks, whereas our approach was based on an aggregate space–time model, and in other contexts (Wakefield, 2008) the effect of aggregation is to distort individual level summaries. A crucial assumption in our calculation is that the infectious period is 1 week, which is clearly an approximation and if we were to allow for a longer duration of infectiveness (e.g. via equation (11)) we would obtain higher estimates of R_0.

Another important issue for data from a passive surveillance system, such as that used to collect the HFMD data, is under-reporting. There are undoubtedly many additional cases in the population who do not visit health centres. Xing *et al.* (2014) carried out a large epidemiological study of HFMD and pointed out the under-reporting; in follow-up letters to Xing *et al.* (2014) (Mao *et al.*, 2014; Leung *et al.*, 2014) the difficulty of adjustment has been highlighted, due to the regional variation in under-reporting. Under-reporting was addressed for HFMD in China by Van Boeckel *et al.* (2016), who had additional detailed birth information. Adapting a previously developed method (Bartlett, 1960), they regressed the cumulative births on the cumulative observed cases, with the reciprocal of the estimated slope giving an estimate of the weekly reporting rate. The observed counts are then inflated by the reciprocal slope for subsequent analysis (with an adjustment for zero observed cases, as described in their supplementary materials). This approach does not account for the uncertainty in the estimation of the rate of under-reporting but could be used with the model that is described here, if we had access to the additional births data.

The neighbourhood contributions to the fit were small (see, for example, Fig. 3). Consequently, in the interests of parsimony, we chose to adopt a relatively simple neighbourhood scheme, not having strata-specific parameters in this component, for example. We examined the robustness to a related scheme in the on-line supplementary materials, section S6. More complex schemes are available (Geilhufe *et al.*, 2014) and could be used within the modelling framework that is described here.

The model framework is very flexible, with different assumptions possible for each of the auto-regressive, neighbourhood and endemic components. This potential is a double-edged sword, as we may in theory place covariates and spatial random effects in all three components. But the decision on where to include these covariates should be made on the basis of knowledge of the particular disease and data collection system, since the aggregate nature of the data limits the information that is available. A separate issue is that caution needs to be taken when developing a covariate model to avoid confounding by location and time. The issue here is that the covariates have spatial and temporal structure and with a complex space–time model, such as that described here, the covariate associations may be distorted, because the spatiotemporal response variability can be attributed to different components of the model. This issue is well known in both the time series (Peng *et al.*, 2006) and the spatial (Wakefield, 2003; Hodges and Reich, 2010; Hughes and Haran, 2013) literatures.

Acknowledgements

This work was supported by National Institutes of Health grant R01 CA095994. The authors are grateful to the China Center for Disease Control for providing the data, to Vladimir Minin for comments on an earlier draft and to Tracy Dong for help with the Stan code. The structure and clarity of the paper were greatly helped by suggestions from the Joint Editor, Associate Editor and two referees.

References

Bartlett, M. S. (1960) The critical community size for measles in the United States. *J. R. Statist. Soc.* A, **123**, 37–44.

Bauer, C., Wakefield, J., Rue, H., Self, S., Feng, Z. and Wang, Y. (2016) Bayesian penalized spline models for the analysis of spatio-temporal count data. *Statist. Med.*, **31**, 1848–1865.

Becker, N. (1989) *Analysis of Infectious Disease Data*. London: Chapman and Hall.

Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1–59.

Bjørnstad, O., Finkenstädt, B. and Grenfell, B. (2002) Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.*, **72**, 169–184.

Bjørnstad, O. and Grenfell, B. (2008) Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environ. Ecol. Statist.*, **15**, 265–277.

Breslow, N. and Clayton, D. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2016) Stan: a probabilistic programming language. *J. Statist. Softwr.*, **76**, 1–32.

Cauchemez, S. and Ferguson, N. (2008) Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *J. R. Soc. Interfc.*, **5**, 885–897.

Chang, H.-L., Chio, C.-P., Su, H.-J., Liao, C.-M., Lin, C.-Y., Shau, W.-Y., Chi, Y.-C., Cheng, Y.-T., Chou, Y.-L., Li, C.-Y., Chen, K.-L. and Chen, K.-T. (2012) The association between enterovirus 71 infections and meteorological parameters in Taiwan. *PLOS One*, **7**, no. 10, article e46845.

Daley, D. and Gani, J. (1999) *Epidemic Modelling: an Introduction*. New York: Cambridge University Press.

Diekmann, O., Heesterbeek, H. and Britton, T. (2012) *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton: Princeton University Press.

Finkenstädt, B. F. and Grenfell, B. T. (2000) Time series modelling of childhood diseases: a dynamical systems approach. *Appl. Statist.*, **49**, 187–205.

Geilhufe, M., Held, L., Skrøvseth, S., Simonsen, G. and Godtliebsen, F. (2014) Power law approximations of movement network data for modeling infectious disease spread. *Biometr. J.*, **56**, 363–382.
Gelman, A. and Hill, J. (2006) *Data Analysis using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
Gibson, G. and Renshaw, E. (1998) Estimating parameters in stochastic compartmental models using Markov chain methods. *Math. Med. Biol.*, **15**, 19–40.
He, D., Ionides, E. and King, A. (2010) Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J. R. Soc. Interfc.*, **7**, 271–283.
Held, L., Höhle, M. and Hofmann, M. (2005) A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statist. Modlling*, **5**, 187–199.
Held, L. and Paul, M. (2012) Modeling seasonality in space-time infectious disease surveillance data. *Biometr. J.*, **54**, 824–843.
Held, L., Schrödle, B. and Rue, H. (2010) Posterior and cross-validatory predictive checks: a comparison of MCMC and INLA. In *Statistical Modeling and Regression Structures—Festschrift in Honour of Ludwig Fahrmeir* (eds T. Kneib and G. Tutz), pp. 91–110. Heidelberg: Physica.
Hodges, J. S. and Reich, B. J. (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *Am. Statistn*, **64**, 325–334.
Höhle, M. (2016) Infectious disease modeling. In *Handbook of Spatial Epidemiology* (eds A. Lawson, S. Banerjee, R. Haining and M. Ugarte), pp. 477–500. Boca Raton: Chapman and Hall–CRC.
Huang, Y., Deng, T., Yu, S., Gu, J., Huang, C., Xiao, G. and Hao, Y. (2013) Effect of meteorological variables on the incidence of hand, foot, and mouth disease in children a: time-series analysis in Guangzhou, China. *BMC Infect. Dis.*, **13**, article 134.
Hughes, J. and Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Statist. Soc.* B, **75**, 139–159.
Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S. and King, A. A. (2015) Inference for dynamic and latent variable models via iterated, perturbed bayes maps. *Proc. Natn. Acad. Sci. USA*, **112**, 719–724.
Knorr-Held, L. and Richardson, S. (2003) A hierarchical model for space–time surveillance data on meningococcal disease incidence. *Appl. Statist.*, **52**, 169–183.
Koepke, A., Longini, Jr, I., Halloran, M., Wakefield, J. and Minin, V. (2016) Predictive modeling of cholera outbreaks in Bangladesh. *Ann. Appl. Statist.*, **10**, 575–595.
Lekone, P. and Finkenstädt, B. (2006) Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, **62**, 1170–1177.
Leroux, B., Lei, X. and Breslow, N. (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment and Clinical Trials* (eds M. Halloran and D. Berry), pp. 179–192. New York: Springer.
Leung, G., Xing, W., Wu, J. and Yu, H. (2014) Hand, foot, and mouth disease in mainland China—authors' reply. *Lancet Infect. Dis.*, **14**, article 1042.
Ma, E., Fung, C., Yip, S., Wong, C., Chunag, S. and Tsang, T. (2011) Estimation of the basic reproduction number of enterovirus 71 and coxsackievirus a16 in hand, foot and mouth disease outbreaks. *Ped. Infect. Dis. J.*, **30**, 675–679.
Mao, Q., Wang, Y. and Liang, Z. (2014) Hand, foot, and mouth disease in mainland China. *Lancet Infect. Dis.*, **14**, article 1041.
McKinley, T., Cook, A. and Deardon, R. (2009) Inference in epidemic models without likelihoods. *Int. J. Biostatist.*, **5**, 24.
Meyer, S. and Held, L. (2014) Power-law models for infectious disease spread. *Ann. Appl. Statist.*, **8**, 1612–1639.
Meyer, S., Held, L. and Höhle, M. (2017) Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *J. Statist. Softwr.*, **77**, 1–55.
Morton, A. and Finkenstädt, B. F. (2005) Discrete time modelling of disease incidence time series by using Markov chain Monte Carlo methods. *Appl. Statist.*, **54**, 575–594.
Mugglin, A., Cressie, N. and Gemmell, I. (2002) Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statist. Med.*, **21**, 2703–2721.
O'Neill, P. and Becker, N. (2001) Inference for an epidemic when susceptibility varies. *Biostatistics*, **2**, 99–108.
O'Neill, P. D. and Roberts, G. O. (1999) Bayesian inference for partially observed stochastic epidemics. *J. R. Statist. Soc.* A, **162**, 121–129.
Paul, M. and Held, L. (2011) Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statist. Med.*, **30**, 1118–1136.
Paul, M., Held, L. and Toschke, A. M. (2008) Multivariate modelling of infectious disease surveillance data. *Statist. Med.*, **27**, 6250–6267.
Peng, R. D., Dominici, F. and Louis, T. A. (2006) Model choice in time series studies of air pollution and mortality. *J. R. Statist. Soc.* A, **169**, 179–203.
Plummer, M. (2008) Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523–539.
Prentice, R., Kalbfleisch, J., Peterson, Jr, A., Flournoy, N., Farewell, V. and Breslow, N. (1978) The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, **71**, 319–392.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.
Spiegelhalter, D., Thomas, A. and Best, N. (1998) *WinBUGS: Bayesian Inference using Gibbs Sampling, Manual v1.2*. Cambridge: Medical Research Council Biostatistics Unit.
Tong, C. and Bible, J. (2009) Global epidemiology of enterovirus 71. *Fut. Virol.*, **4**, 501–510.
Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M. (2010) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interfc.*, **6**, 187–202.
Van Boeckel, T., Takahashi, S., Liao, Q., Xing, W., Lai, S., Hsiao, V., Liu, F., Zheng, Y., Chang, Z., Yuan, C., Metcalf, C. J., Yu, H. and Grenfell, B. T. (2016) Hand, foot, and mouth disease in China: critical community size and spatial vaccination strategies. *Scient. Rep.*, **6**, article 25248.
Wakefield, J. (2008) Ecologic studies revisited. *A. Rev. Publ. Hlth*, **29**, 75–90.
Wakefield, J. (2013) *Bayesian and Frequentist Regression Methods*. New York: Springer.
Wakefield, J. C. (2003) Sensitivity analyses for ecological regression. *Biometrics*, **59**, 9–17.
Wakefield, J., Dong, T. and Minin, V. (2018) Spatio-temporal analysis of surveillance data. In *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, P. Guttorp and M. Fuentes). Boca Raton: CRC Press.
Wang, Y., Feng, Z., Yang, Y., Self, S., Gao, Y., Longini, I., Wakefield, J., Zhang, J., Wang, L., Chen, X., Yao, L., Stanaway, J., Wang, Z. and Yang, W. (2011) Hand, foot and mouth disease in China: patterns and spread and transmissibility. *Epidemiology*, **22**, 781–792.
Wu, H., Wang, H., Wang, Q., Xin, Q. and Lin, H. (2014) The effect of meteorological factors on adolescent hand, foot, and mouth disease and associated effect modifiers. *Globl Hlth Actn*, **7**, article 24664.
Xing, W., Liao, Q., Viboud, C., Zhang, J., Sun, J., Wu, J., Chang, Z., Liu, F., Fang, V., Zheng, Y., Cowling, B., Varma, J., Farrar, J., Leung, G. and Yu, H. (2014) Hand, foot, and mouth disease in China, 2008–12: an epidemiological study. *Lancet Infect. Dis.*, **14**, 308–318.
Yamin, D., Jones, F. K., DeVincenzo, J. P., Gertler, S., Kobiler, O., Townsend, J. P. and Galvani, A. P. (2016) Vaccination strategies against respiratory syncytial virus. *Proc. Natn. Acad. Sci. USA*, **113**, 13239–13244.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary materials for the stratified space-time infectious disease modeling: with an application to hand, foot and mouth disease in China'.