STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:6250–6267
Published online 17 September 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3440

Multivariate modelling of infectious disease surveillance data

M. Paul[1], L. Held[1,*,†] and A. M. Toschke[2]

[1]*Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Switzerland*
[2]*Department of Public Health Sciences, Division of Health and Social Care Research, King's College London, U.K.*

SUMMARY

This paper describes a model-based approach to analyse multivariate time series data on counts of infectious diseases. It extends a method previously described in the literature to deal with possible dependence between disease counts from different pathogens. In a spatio-temporal context it is proposed to include additional information on global dispersal of the pathogen in the model. Two examples are given: the first describes an analysis of weekly influenza and meningococcal disease counts from Germany. The second gives an analysis of the spatio-temporal spread of influenza in the U.S.A., 1996–2006, using air traffic information. Maximum likelihood estimates in this non-standard model class are obtained using general optimization routines, which are integrated in the R package `surveillance`. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: infectious disease surveillance; multivariate time series of counts; space–time models

1. INTRODUCTION

A major challenge in infectious disease epidemiology remains the analysis of data on notifiable diseases typically collected by national surveillance systems. Time series on counts of infectious diseases often show a regular pattern over time such as long-term trends or seasonality but also occasional outbreaks. This mixture of endemic and epidemic behaviours has to be taken into account when modelling such data. Another characteristic is overdispersion with respect to the usual Poisson assumption. Besides, issues such as under-reporting or reporting delays are quite common.

The probably best-studied stochastic models for the spread of an infectious disease over time are mechanistic models such as the chain-binomial model and related continuous time models such as the SIR model [1, 2]. These models directly describe the infection process of the spread from

*Correspondence to: L. Held, Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland.
†E-mail: leonhard.held@ifspm.uzh.ch

Contract/grant sponsor: Swiss National Science Foundation (SNF)

Received 25 March 2008
Accepted 8 August 2008

person-to-person on an individual level. A key quantity is the basic reproduction number, which together with the number of infectious and susceptibles, allows to provide answers, e.g. about the size and duration of outbreaks or the effect of vaccination programs.

However, a major requirement of epidemic models is that the epidemic process is completely observed. In particular, one has to know at each time point the number of infected and the number of susceptible individuals. If information about susceptibles is not available, the chain-binomial and SIR models can be approximated by a branching process where one assumes an unlimited amount of susceptibles [2]. In large populations such an approximation is especially good since the depletion of susceptibles in the population is, at least at initial stages of an outbreak, negligible. Provided that the proportion of infected individuals stays small relative to the number of susceptibles, the branching process approximation will continue to hold throughout the course of an outbreak [3, 4].

In a surveillance setting, the available data are often spatially and temporally aggregated and information about susceptibles is rarely available. For example, Finkenstädt *et al.* [5, 6] reconstruct the number of susceptibles using data on cases and on births in order to explain extinction and recurrence of epidemics observed in measles. In general, pure mechanistic modelling is too ambitious for routinely collected surveillance data and alternative approaches have been developed.

Purely empirical models, such as log-linear Poisson regression, are not able to capture epidemic outbreaks adequately, so several extensions of Poisson regression models have been suggested. For instance, Zeger and Qaqish [7] propose a Markov regression model with a multiplicative effect of past observations on the disease incidence. For multivariate time series, Knorr-Held and Richardson [8] describe a model where previous counts also enter multiplicatively, modulated by latent binary indicators, which are assumed to follow a two-stage hidden Markov model.

In contrast, Held *et al.* [9] proposed a model based on a branching process with immigration and Poisson offspring. By construction, previous counts now enter additively rather than multiplicatively. The model is extended to allow for seasonality and overdispersion and can be estimated with the maximum likelihood (ML) techniques. This approach yielded promising results in the analysis of univariate and multivariate time series on a specific disease caused by a single pathogen.

However, interdependencies between diseases caused by different pathogens might particularly be of interest to further understand the dynamics of such diseases. For example, viral infections may cause physical damage to respiratory cells and thus facilitate bacterial adherence, clearing the way for bacterial disease [10]. Several studies give both clinical and epidemiological support that influenza infections predispose meningococcal disease [11–14]. For Illustration, Figure 1 shows the weekly number of influenza (labelled as FLU) and meningococcal disease cases (labelled as MEN) in Germany, 2001–2006, obtained from the German national surveillance system for notifiable diseases, administered by the Robert Koch Institute (RKI) [15]. The influenza counts show yearly outbreaks of different severity during the winter. The meningococcal disease counts also display a seasonal pattern with small outbreaks during the winters of 2003 and 2005, which seem to coincide with the two biggest outbreaks of influenza.

Another possible scenario is that several infections are transmitted via the same route. If there is an underlying increase in transmission, e.g. due to varying contact rates, this will induce a directionless correlation between the diseases considered. For example, Farrington *et al.* [16] analyse data on several airborne infections while De Angelis *et al.* [17] consider several sexually transmitted diseases.

In this paper we extend the multivariate model introduced in Held *et al.* [9] to analyse data from different pathogens. In particular, overdispersion as well as seasonality is allowed to vary across diseases and additional parameters capturing a directed influence from one disease to the other are

DOI: 10.1002/sim

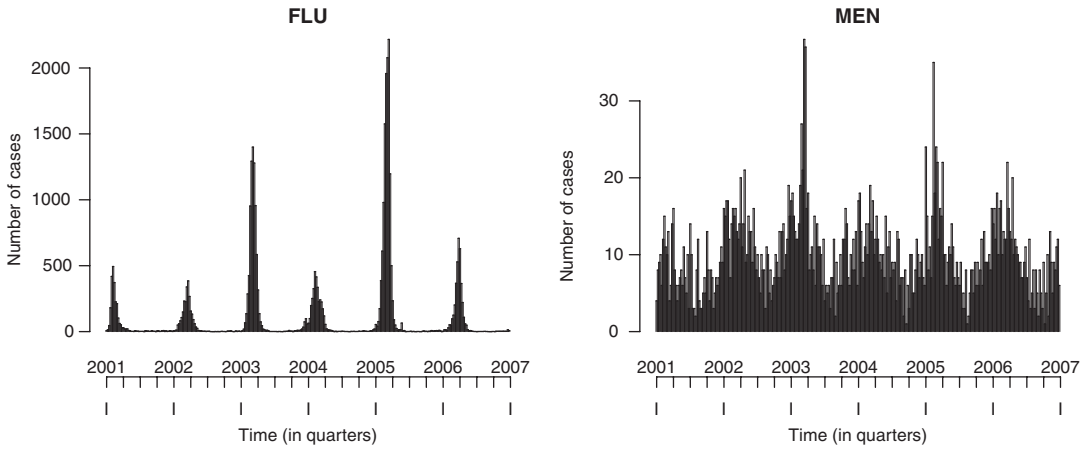

Figure 1. Weekly number of influenza (FLU) and meningococcal disease (MEN) cases in Germany, 01/2001–52/2006.

introduced as described in Section 2. In Section 3.1 we use this model framework to investigate a possible directed association between influenza and meningococcal disease cases for the data described above. Furthermore, this analysis gives empirical evidence of the better fit of additive rather than multiplicative models.

The multivariate branching process model proposed in Held *et al.* [9] can also be applied to spatio-temporal surveillance data. In Section 2.1, we describe a further extension of the model by including additional weights. External data such as information on travel intensities between spatial units can now be incorporated. As an example, in Section 3.2 we consider the weekly number of deaths from influenza and pneumonia in the U.S.A., previously analysed in Brownstein *et al.* [18]. A multivariate model including air traffic information provides a better fit than simple adjacency-based models or models without interdependencies between geographical regions. We close with some discussion in Section 4.

2. THE MODEL FRAMEWORK

To begin, let $y_{i,t}$ denote the number of cases observed in 'unit' i at time t, $i=1,\ldots,m$, $t=1,\ldots,T$. A unit might represent not only a single disease observed, e.g. in several geographical regions or in different age groups, but also different pathogens observed in one location. We might, for example, be interested in multiple diseases transmitted via the same route, e.g. airborne infections. A simple version of the model considered assumes that the counts are negative binomial distributed, $y_{i,t}|y_{i,t-1}\sim\text{NegBin}(\mu_{i,t},\psi)$, with conditional mean

$$\mu_{i,t}=\lambda y_{i,t-1}+\exp(\eta_{i,t}) \tag{1}$$

and conditional variance

$$\mu_{i,t}(1+\psi\mu_{i,t})$$

DOI: 10.1002/sim

here $\psi>0$ is an additional overdispersion parameter, see below for details. The disease incidence $\mu_{i,t}$ can thus be decomposed additively into two parts. Following Held *et al.* [9] we call the first part $\zeta_{i,t}=\lambda y_{i,t-1}$ (where λ is an unknown autoregressive parameter) the 'epidemic' component, and the second part $\nu_{i,t}=\exp(\eta_{i,t})$ the 'endemic' component. In Section 2.2 we describe parametric approaches to model $\nu_{i,t}$.

The epidemic component should be able to capture occasional outbreaks whereas the endemic component explains a baseline rate of cases that is persistent with a stable temporal pattern. For example, it is quite common to distinguish 'sporadic' and epidemic outbreaks in the incidence analysis of meningitis [8].

The negative binomial model allows for possible overdispersion due to, for example, under-reporting or unobserved covariates that affect the disease incidence. For $\psi=0$ the negative binomial model reduces to a Poisson model where the conditional variance is equal to the conditional mean $\mu_{i,t}$. Note that Held *et al.* [9] use $1/\psi$ as the overdispersion parameter.

In the above model, overdispersion is identical in every unit. If the units are age groups or regions, this may be a realistic assumption. However, when the units correspond to different types of diseases this assumption is unlikely to hold and unit-specific overdispersion parameters ψ_i may be used instead.

2.1. Epidemic component

The epidemic component in (1) is modelled by an autoregression on the number of cases $y_{i,t-1}$ in unit i at the previous time point $t-1$. The inclusion of previous cases allows for temporal dependence beyond seasonal patterns within a unit. However, the model will not be able to explain the spread of a disease across units. Hence, Held *et al.* [9] suggest to include the sum of the previous number of cases $y_{j,t-1}$ in other units $j\neq i$ as a potential explanatory variable for the disease incidence in unit i. Depending on the context, the other units may be e.g. all other units or solely geographically neighbouring units.

Here, we consider a more general version of the epidemic component

$$\zeta_{i,t}=\lambda_i y_{i,t-1}+\phi_i \sum_{j\neq i} w_{ji} y_{j,t-l} \tag{2}$$

where $y_{j,t-l}$ denotes the number of cases observed in unit j at time $t-l$ with lag $l\in\{1,2,\ldots\}$, and w_{ji} are suitably chosen weights. The simplest choice for the weights is $w_{ji}=1$ for all $j\neq i$, more elaborate choices will be discussed in the following.

The additional autoregressive parameters ϕ_i quantify the influence of $y_{j,t-l}$, $j\neq i$, on $y_{i,t}$. Unit-specific autoregressive parameters are useful if the units correspond to different types of diseases. Depending on the duration of the incubation and infectious period it might also be necessary to look at lagged counts $y_{j,t-l}$ with lag $l>1$. For example, consider both influenza and meningococcal infections. There is experimental as well as epidemiological evidence that respiratory viral infections predispose for bacterial disease [10]. Studies showed that patients with severe meningococcal disease were more likely than control subjects to show serological evidence of recent influenza infection [11, 14]. A possible directed association between influenza and subsequent meningococcal disease in Germany will be analysed in Section 3.1.

Suppose surveillance data on the same pathogen are available for several geographic locations $i=1,\ldots,m$. Possible choices for the weights are then, e.g. $w_{ji}=\mathbb{1}(j\sim i)$, where $\mathbb{1}$ is the indicator function and $j\sim i$ denotes all units that are adjacent to i. Then only regions adjacent to region i

DOI: 10.1002/sim

are taken into account and the sum of counts in adjacent regions enters as an explanatory variable. Such binary weights have been used in Held *et al.* [9] to analyse the weekly number of measles cases in the districts of Lower Saxony, Germany.

However, perhaps more natural is $w_{ji}=\mathbb{1}(j\sim i)/|k\sim j|$, where $|k\sim j|$ denotes the number of neighbours of region j. Thereby we assume that a proportion of infected individuals, say 80 per cent, stays in region j thus being able to infect other individuals in this region. The remaining 20 per cent spread out uniformly among adjacent regions, i.e. individuals in a certain region i (adjacent to j) can also be infected by $20/|k\sim j|$ per cent of cases introduced from region j. This assumption is made for all regions.

The dispersal of cases in space is not necessarily only local but also global. Linking of parallel time series based on adjacencies may then be unrealistic. For instance, SIR-type models on a local level (i.e. cities) have been combined with air transportation data to model the spatio-temporal spread of infectious diseases such as influenza and SARS [19–21]. An alternative choice would be to include travel information in the weights w_{ji} if such information is available. In Section 3.2 we will use the number of airline passengers obtained from the TranStats database, U.S. Department of Transportation [22], to analyse influenza mortality in the U.S.A.

2.2. *Endemic component*

The endemic component includes terms to describe differences between units and seasonality and is specified as

$$\log(\nu_{i,t})=\eta_{i,t}=\alpha_i+\sum_{s=1}^{S}(\gamma_s\sin(\omega_s t)+\delta_s\cos(\omega_s t)) \tag{3}$$

where S is the number of harmonics to include and ω_s are Fourier frequencies, e.g. $\omega_s=2\pi s/52$ for weekly data. Modelling seasonal variation of infectious diseases through superposition of harmonic waves goes back to Serfling [23] and has since then been used in various models (e.g. [13, 24]). An alternative representation of the seasonal terms in (3) as

$$\gamma_s\sin(\omega_s t)+\delta_s\cos(\omega_s t)=A_s\sin(\omega_s t+\varphi_s)$$

with $A_s=\sqrt{\gamma_s^2+\delta_s^2}$ and $\tan(\varphi_s)=\delta_s/\gamma_s$ is easier to interpret. Thus, the parameter A_s marks the amplitude of the seasonal component s, while φ_s represents the phase difference [25].

The parameter α_i in (3) allows for different incidence levels in each of the m units (age groups, regions or pathogens). This assumption is reasonable since, e.g. the willingness of a sick person to seek medical advice might depend on the age of that person, for example if case severity does depend on age. Also, the compliance of general practitioners or local public health offices to report cases might differ by area. A further extension of the model is to also let the seasonality terms vary across units, i.e.

$$\log(\nu_{i,t})=\eta_{i,t}=\alpha_i+\sum_{s=1}^{S_i}(\gamma_{i,s}\sin(\omega_s t)+\delta_{i,s}\cos(\omega_s t)) \tag{4}$$

Note that the number of harmonic waves S_i might also depend on unit i. If we are interested in modelling related diseases, pathogen-specific seasonality is a sensible assumption.

An additional modification is to consider different (standardized) population sizes $n_{i,t}$ by assuming $\mu_{i,t}=\zeta_{i,t}+n_{i,t}\nu_{i,t}$. Time-dependent population adjustment $n_{i,t}$ might be useful for

DOI: 10.1002/sim

subgroups (age, geographical region, etc.) with a varying distribution over time. However, relatively long time series (covering several years up to decades, say) are typically needed for a pronounced change in population proportions. Adjustment for different but constant proportions among units, n_i, is not necessary since this will be completely absorbed by the unit-specific incidence levels α_i.

2.3. Likelihood inference

Conditional on $y_{i,t-1}, \ldots, y_{i,t-l}$, $i = 1, \ldots, m$, the counts $y_{i,t}$ are assumed to be negative binomial distributed with mean

$$\mu_{i,t}(\boldsymbol{\theta}_i) = \mu_{i,t} = \zeta_{i,t} + \nu_{i,t} \tag{5}$$

where $\zeta_{i,t}$ is given in (2), $\nu_{i,t}$ is given in (4) and $\boldsymbol{\theta}_i = (\lambda_i, \phi_i, \alpha_i, \gamma_{i,1}, \ldots, \gamma_{i,S_i}, \delta_{i,1}, \ldots, \delta_{i,S_i})^{\mathrm{T}}$. Those parameters that are not specified are omitted. With $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m, \psi_1, \ldots, \psi_m)^{\mathrm{T}}$ the log-likelihood is given as

$$l(\boldsymbol{\theta}) = \sum_{i,t} l_{i,t}(\boldsymbol{\theta}_i, \psi_i)$$

where

$$l_{i,t}(\boldsymbol{\theta}_i, \psi_i) \propto \log\Gamma\left(y_{i,t} + \frac{1}{\psi_i}\right) - \log\Gamma\left(\frac{1}{\psi_i}\right) + \frac{1}{\psi_i}\log\left(\frac{1}{1+\psi_i\mu_{i,t}(\boldsymbol{\theta}_i)}\right) + y_{i,t}\log\left(\frac{\psi_i\mu_{i,t}(\boldsymbol{\theta}_i)}{1+\psi_i\mu_{i,t}(\boldsymbol{\theta}_i)}\right)$$

is the log-likelihood contribution of observation $y_{i,t}$ and $\Gamma(\cdot)$ is the gamma function [26, p. 255].

If $\lambda = 0$ and for fixed ψ, model (1) corresponds to a log-linear generalized linear model (GLM) and can be fitted with standard software for GLMs to obtain the ML estimates $\hat{\boldsymbol{\theta}}$ [27]. Otherwise, the log-likelihood $l(\boldsymbol{\theta})$ needs to be optimized numerically using generic optimization routines such as the quasi-Newton method BFGS implemented in the R function `optim`. Held *et al.* [9] used numerical approximations of the score function and Fisher information matrix to fit the model, whereas we use analytical derivatives to speed up the computation of the ML estimates. The fitting procedure is implemented in the R package `surveillance` [28] available from the Comprehensive R Archive Network at `http://cran.r-project.org`. Implementation details can be found in Appendix A.

3. APPLICATION TO DATA

3.1. Influenza and meningococcal disease

Associations between influenza and subsequent rates and severities of meningococcal disease have been well documented in the literature (see e.g. References [11–14] or the reviews of Hament *et al.* [10], Brundage [29] and references therein). We analysed the influence of influenza on invasive meningococcal disease based on the weekly number of cases of both disease types in Germany, 2001–2006 as discussed in Section 1 and shown in Figure 1.

DOI: 10.1002/sim

Table I. Univariate analysis of meningococcal infections in Germany, 01/2001–52/2006.

S	$\hat{\lambda}$ (s.e.)	$\hat{\psi}$ (s.e.)	$\log L$	p	AIC
1	—	0 (fixed)	−872.1	3	1750.2
1	—	0.06 (0.01)	−850.2	4	1708.3
1	0.16 (0.06)	0.05 (0.01)	−845.6	5	1701.2
2	0.16 (0.06)	0.05 (0.01)	−845.5	7	1705.0

The log-likelihood is denoted by $\log L$, p is the number of parameters and AIC $=-2\log L+2p$.

Note that the influenza data suffer from under-reporting: the observed counts comprise only laboratory confirmed cases and thus, only a fraction of all influenza cases is actually recorded since the illness caused by influenza is often too slight to warrant medical attention. Nevertheless, the data are still able to reflect the temporal course of the disease. On the other hand, a capture–recapture-analysis showed that the degree of ascertainment for meningococcal disease (reported to the RKI) is quite high [30].

Table I summarizes results from a univariate analysis of the meningococcal disease data alone. Based on $S=1$ seasonal term, a negative binomial model instead of the Poisson model with fixed $\psi=0$ results in a significant increase of the maximized log-likelihood from -872.1 to -850.2. Additional inclusion of the autoregressive parameter λ leads to a further substantial improvement of the maximized log-likelihood. The ML estimate of λ is 0.16 (0.06) indicating a weak dependence on the number of cases in the previous week after adjustments for seasonal effects. Higher degrees S for seasonality give only slight improvements of the maximized log-likelihood, so the best model according to the model choice criterion AIC is the negative binomial model with $S=1$ seasonal term and the autoregressive parameter λ, see Table I.

Based on a similar univariate analysis of the influenza data, the best model according to AIC includes the overdispersion parameter, an autoregressive term and $S=3$ seasonal terms.

Table II now summarizes the results of selected multivariate models. The conditional mean of the most general formulation is specified as

$$\begin{pmatrix}\mu_{\text{men},t}\\ \mu_{\text{flu},t}\end{pmatrix}=\begin{pmatrix}\lambda_{\text{men}} & \phi_{\text{men}}\\ \phi_{\text{flu}} & \lambda_{\text{flu}}\end{pmatrix}\begin{pmatrix}\text{MEN}_{t-1}\\ \text{FLU}_{t-1}\end{pmatrix}+\begin{pmatrix}\nu_{\text{men},t}\\ \nu_{\text{flu},t}\end{pmatrix}$$

The first model shown does not allow for any interdependencies between influenza and meningococcal disease (i.e. $\phi_{\text{men}}=\phi_{\text{flu}}=0$) and is based on the respective best univariate models with $\log L$ being the sum of the log-likelihood values from the univariate analyses. The AIC for this model is 3807.5. The second model, which includes an influence from influenza on meningococcal disease (denoted by 'flu → men' in Table II), shows a better fit than the model without interaction (AIC $=3791.9$). We also looked at the 'reverse' model that includes an influence from meningococcal disease on influenza ('men → flu'). Since the inclusion of past meningococcal disease cases plays no role, $\hat{\phi}_{\text{flu}}\approx 0$, this supports that the association between meningococcal disease and influenza is directed. Finally, we also fitted a model that includes both 'flu → men' and 'men → flu'. Results for this model are virtually identical to those obtained based on the second model without the additional 'men → flu' term. Indeed, the additional parameter does not improve the log-likelihood $\log L$, therefore AIC is increased by 2 to AIC $=3793.9$.

DOI: 10.1002/sim

Table II. Multivariate analysis of influenza and meningococcal disease in Germany, 01/2001–52/2006.

$\hat{\lambda}$ (s.e.)		$\hat{\phi}$ (s.e.)		$\hat{\psi}$ (s.e.)			
flu	men	men → flu	flu → men	flu	men	log L	p
0.74 (0.05)	0.16 (0.06)	—	—	0.29 (0.04)	0.05 (0.01)	−1889.7	14
0.74 (0.05)	0.10 (0.06)	—	0.005 (0.001)	0.29 (0.04)	0.04 (0.01)	−1881.0	15
0.74 (0.05)	0.16 (0.06)	4e−07 (1e−04)	—	0.29 (0.04)	0.05 (0.01)	−1889.7	15
0.74 (0.05)	0.10 (0.06)	4e−07 (1e−04)	0.005 (0.001)	0.29 (0.04)	0.04 (0.01)	−1881.0	16

The endemic component $\nu_{i,t}$ includes $S=3$ seasonal terms for the FLU data and $S=1$ seasonal term for the MEN data. The log-likelihood is denoted by $\log L$ and p is the number of parameters.

Figure 2 shows the observed influenza and meningococcal disease cases together with the fitted means $\hat{\mu}_{\text{flu},t}$ and $\hat{\mu}_{\text{men},t}$ for the model with interaction 'flu → men'. The means are separated into two, respectively, three additive components: an endemic and an autoregressive component for the FLU data and an endemic, an autoregressive and an 'influenza-driven' component for the MEN data. The two outbreaks of meningococcal disease in 2003 and 2005 are explained primarily by the influence of influenza, which indicates an association between influenza cases and subsequent meningococcal disease cases. Overall, the model gives a quite good fit to the data and is able to explain the seasonality and the outbreaks in the meningococcal disease data, as well as the different severity of the influenza outbreaks.

Deviance residuals (see e.g. [31]) and corresponding autocorrelation functions are also shown in Figure 2. The residuals seem to be roughly uncorrelated, perhaps with the exception of the lag two autocorrelation for influenza. We also looked at the cumulative periodogram of the residual series to assess compatibility with white noise: no obvious deviation from white noise could be seen. Such tests for white noise, however, neglect the effects of parameter estimation and are therefore only indications and not strictly valid [32].

So far, only the number of influenza cases in week $t-1$ (lag $l=1$) has been considered as explanatory variable for the meningococcal disease incidence in week t. To further investigate the relationship between influenza and meningococcal disease we also considered a delayed effect of influenza towards meningococcal disease of up to three weeks. Table III lists the ML estimates of ϕ_{men} for several negative binomial models with conditional mean for the meningococcal disease data specified as

$$\mu_{\text{men},t} = \lambda_{\text{men}}\text{MEN}_{t-1} + \phi_{\text{men}}\text{FLU}_{t-l} + \nu_{\text{men},t}$$

where the endemic component $\nu_{\text{men},t}$ includes $S=1$ seasonal term, and lags $l \in \{-3, \ldots, 3\}$. Forward lags were also included in the analysis to assess whether the association is unidirectional in time (compare with the analysis in [12]). If there is a directed association the estimated values for ϕ_{men} should be asymmetric around lag zero. As in Hubert *et al.* [12] there is still a positive association for negative lags, but the estimated effect is largest for lag one and slightly smaller for lag zero. This indicates that changes in the incidence rate of influenza might be associated with one-week delayed ($l=1$) or with simultaneous ($l=0$) changes in the incidence rate of meningococcal disease although the evidence is not very strong. Since we only know the week when a case was reported and do not have the exact dates of infection, some event times might be misclassified, which would weaken the autocorrelation and disturb the lag.

For comparison, we also fitted a 'multiplicative' autoregressive conditional model as suggested in Zeger and Qaqish [7] to the meningococcal disease data. In their model, the deviation of the

DOI: 10.1002/sim

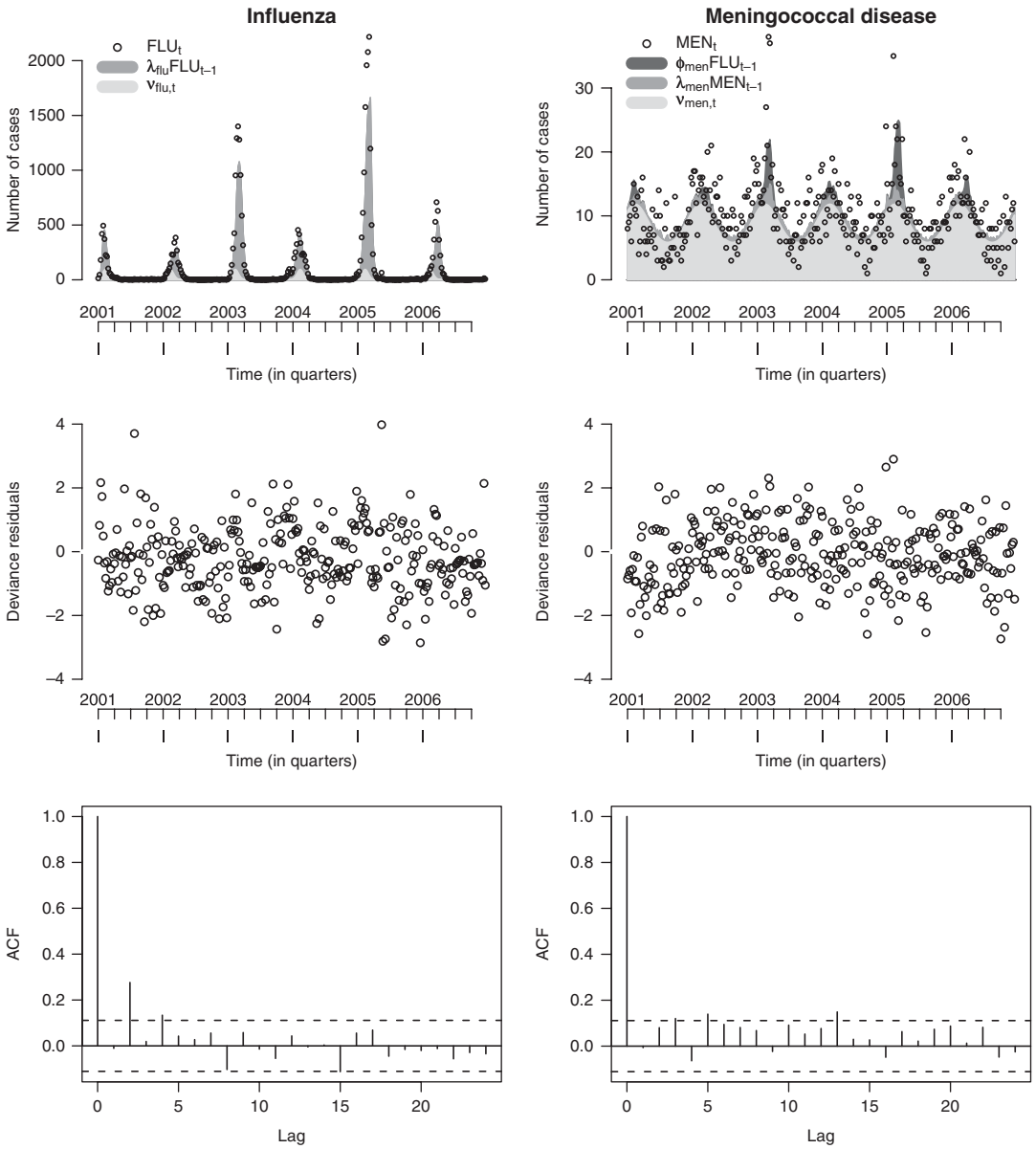

Figure 2. Observed number of influenza and meningococcal disease cases with fitted mean, deviance residuals and corresponding autocorrelation function of the model with interaction 'flu → men'.

logarithm of the observed counts y_{t-1} at time $t-1$ from the linear predictor η_{t-1} at time $t-1$ enters as an explanatory variable:

$$\log(\mu_t) = \eta_t + \theta(\log(y^*_{t-1}) - \eta_{t-1})$$

DOI: 10.1002/sim

Table III. Analysis of meningococcal disease data with several lagged influenza counts as explanatory variables.

lag l	$\hat{\phi}\times 10^3$ (s.e. $\times 10^3$)
3	2.92 (1.30)
2	4.54 (1.41)
1	5.32 (1.42)
0	5.30 (1.39)
−1	4.68 (1.31)
−2	3.73 (1.26)
−3	2.30 (1.22)

The mean is specified as $\mu_{\text{men},t}=\lambda_{\text{men}}\text{MEN}_{t-1}+\phi_{\text{men}}\text{FLU}_{t-l}+\nu_{\text{men},t}$, where the endemic component $\nu_{\text{men},t}$ includes $S=1$ seasonal term.

To avoid non-existence of the logarithm, any zero values of y_{t-1} are replaced by a constant c, $0<c<1$, i.e. $y^*_{t-1}=\max\{y_{t-1},c\}$. Note that past cases y_{t-1} are not simply added directly to the linear predictor η_t as would seem natural, because such a model cannot describe positive association without growing exponentially in time [33, Section 10.4]. The conditional mean μ_t of $y_t|y_{t-1}$ is thus given by

$$\mu_t=\exp(\eta_t)\left[\frac{y^*_{t-1}}{\exp(\eta_{t-1})}\right]^{\theta} \tag{6}$$

so past counts y^*_{t-1} act multiplicatively on the conditional mean μ_t relative to $\exp(\eta_{t-1})$. Extensions and alternative forms of this autoregressive conditional model have been suggested elsewhere [34, 35].

The exact form of the conditional mean μ_t for the meningococcal disease data can be found in Table IV. Throughout, a negative binomial observation model is used and the linear predictor η_t always contains $S=1$ seasonal term. For the multiplicative model (6) any zero values of MEN_t or FLU_t are replaced with $c=0.1$. All AIC values given in Table IV are computed using the log-likelihood contributions of only the MEN data. Letting previous counts of both meningococcal disease and influenza act multiplicatively instead of additively on the mean μ_t gives a worse fit according to AIC. Besides, the introduction of an arbitrary constant c to avoid non-existence of the logarithm is not necessary for the additive model and the parameters are easier to interpret.

3.2. Influenza in the U.S.A.

As part of its national influenza surveillance effort, the Centers for Disease Control and Prevention (CDC) receive weekly mortality reports from 122 cities and metropolitan areas in the U.S.A. within 2–3 weeks from the date of death. These reports summarize the total number of deaths occurring in these cities/areas each week, as well as the number due to pneumonia and influenza. Figure 3 shows the weekly number of deaths from influenza and pneumonia obtained from the CDC 121 Cities Mortality Reporting System for weeks 40/1996 to 39/2006 in nine major geographic regions of the U.S.A. [36]. A map of these regions is shown in Figure 4.

These data have been analysed in Brownstein *et al.* [18]. The authors studied the influence of long-range airline travel on the inter-regional influenza spread and found empirical evidence that

DOI: 10.1002/sim

Table IV. Comparison of 'additive' and 'multiplicative' models for the meningococcal disease data.

Additive model	AIC	Multiplicative model	AIC
$\mu_t = \exp(\eta_t) + \lambda \text{MEN}_{t-1}$	1701.2	$\mu_t = \exp(\eta_t) \left[\frac{\text{MEN}^*_{t-1}}{\exp(\eta_{t-1})} \right]^{\theta_1}$	1703.3
$\mu_t = \exp(\eta_t) + \lambda \text{MEN}_{t-1} + \phi \text{FLU}_{t-1}$	1685.7	$\mu_t = \exp(\eta_t + \theta_2 \log(\text{FLU}^*_{t-1})) \left[\frac{\text{MEN}^*_{t-1}}{\exp(\tilde{\eta}_{t-1})} \right]^{\theta_1}$	1700.4
$\mu_t = \exp(\eta_t) + \phi \text{FLU}_{t-1}$	1686.5	$\mu_t = \exp(\eta_t + \theta_2 \log(\text{FLU}^*_{t-1}))$	1704.9

The endemic component $\nu_t = \exp(\eta_t)$ includes $S=1$ seasonal term, $\text{MEN}^*_t = \max\{\text{MEN}_t, 0.1\}$, $\text{FLU}^*_t = \max\{\text{FLU}_t, 0.1\}$ and $\tilde{\eta}_{t-1} = \eta_{t-1} + \theta_2 \log(\text{FLU}^*_{t-2})$.

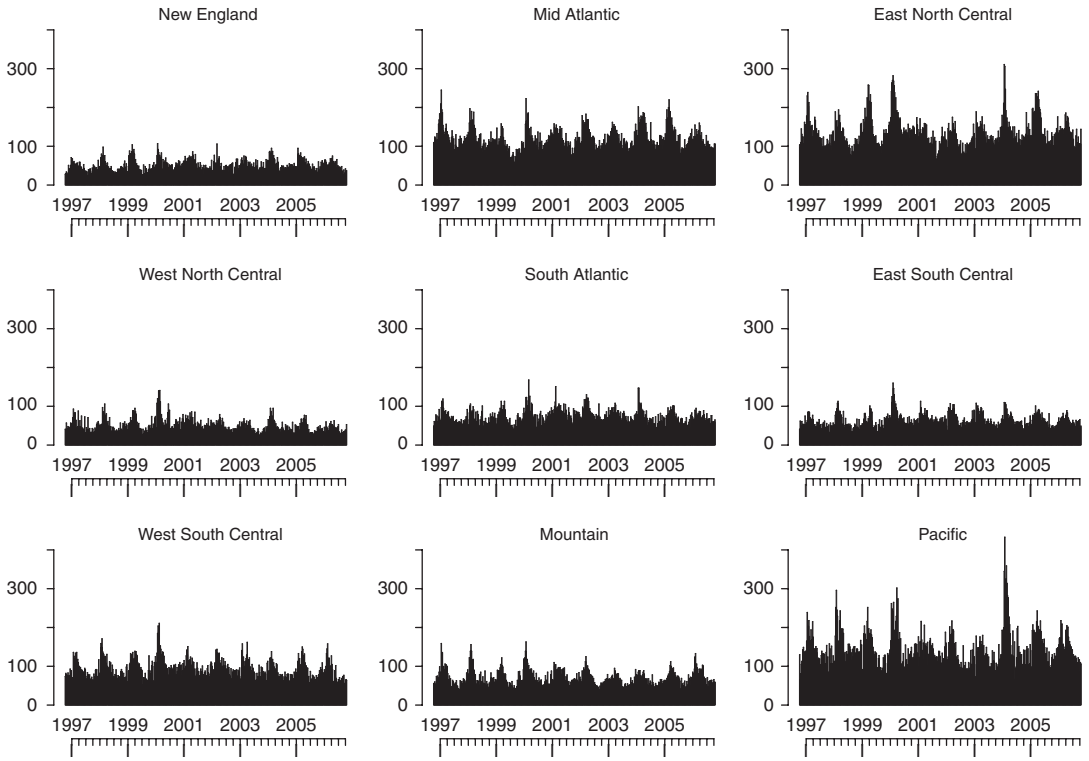

Figure 3. Weekly number of deaths from influenza and pneumonia in the U.S.A. 40/1996–39/2006.

air travel plays a role in the annual spread of influenza in the U.S.A. We applied our model to these data and compared several weights w_{ji} in the epidemic component: weights based on geographic adjacencies and weights including air travel information.

With regard to the form of seasonality $S=4$ seasonal terms are included in the endemic component $\nu_{i,t}$. The epidemic components includes an autoregressive parameter λ as well as region-specific parameters ϕ_i, $i=1,\ldots,9$. As weights in (2) we use firstly $w_{ji} = \mathbb{1}(j \sim i)$, i.e. only neighbouring

DOI: 10.1002/sim

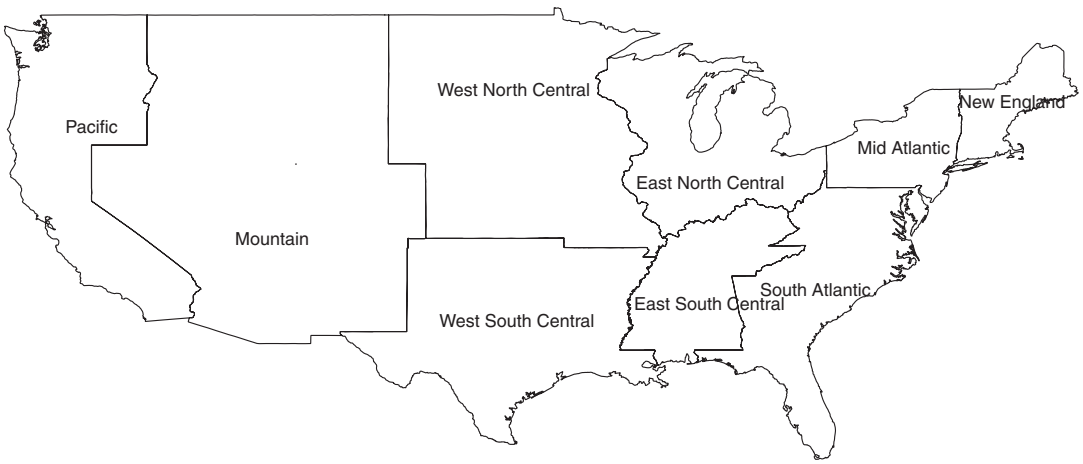

Figure 4. Map of nine major geographical regions of the U.S.A. as defined by the CDC.

regions are considered. Two regions are defined to be neighbours if they share a common border. Secondly, we use geographic weights $\mathbb{1}(j\sim i)/|k\sim j|$ as discussed in Section 2.1. Finally, we use average numbers of passengers travelling by air from region j to region i relative to the population in region j. Data on the yearly number of passengers travelling by air within the U.S.A. were obtained from the TranStats database, U.S. Department of Transportation [22] and data on state population estimates were obtained from the Population Estimates Program, U.S. Bureau of the Census [37]. Denote n_i the population in region i (in the year 2000) and

$$p_{ji} = \#\text{ passengers per week from } j \to i \quad \text{(average over years 1996–2006)}$$

$$p_{ji}(\text{yearly}) = \#\text{ passengers per week from } j \to i \quad \text{(average per year)}$$

That means that the weights p_{ji}/n_j are constant for all time points $t=1,\ldots,522$, whereas the weights $p_{ji}(\text{yearly})/n_j$ change each year.

Results of these models are summarized in Table V. Note that only the smallest and largest value of the $\hat{\phi}_i$'s are given in the table instead of all nine values. It can be seen that the inclusion of both the autoregressive parameters λ and ϕ_i improves the fit according to AIC. There is not much difference in terms of maximized log-likelihood between the two models that use geographic weights, the second model that includes the number of neighbours $|k\sim j|$ performing slightly better. Both models that include travel information perform better than the models with geographic weights, the model using time-varying weights $p_{ji}(\text{yearly})/n_j$ being the best.

The actual value of ϕ_i depends on the magnitude of $\sum_j w_{ji}y_{j,t-1}$, therefore a direct comparison of these values from different models is not useful. However, one can look at another quantity. The general model can also be written in a multivariate fashion as

$$\boldsymbol{\mu}_t = \boldsymbol{\Lambda}\mathbf{y}_{t-1} + \boldsymbol{\nu}_t \tag{7}$$

where $\boldsymbol{\mu}_t$, $\mathbf{y}_{t-1}$ and $\boldsymbol{\nu}_t$ are vectors of length m and $\boldsymbol{\Lambda}$ is a $m\times m$ matrix with elements λ_i on the diagonal and elements $(\Lambda)_{ij}=\phi_i w_{ji}$ for $i\neq j$. For constant $\boldsymbol{\nu}_t$ model (7) corresponds to a multivariate branching process with immigration. If the largest eigenvalue of $\boldsymbol{\Lambda}$ is smaller than

DOI: 10.1002/sim

Table V. Multivariate analysis of influenza mortality in the U.S.A.

w_{ji}	$\hat{\lambda}$ (s.e.)	$\hat{\phi}_i$ (s.e.)	$\hat{\psi}$ (s.e.)	AIC	max EV
—	—	—	0.04 (0.001)	40300.5	—
—	0.34 (0.01)	—	0.03 (0.001)	39693.6	0.34
$\mathbb{1}(j \sim i)$	0.30 (0.01)	0.01 (0.01)–0.23 (0.08)	0.03 (0.001)	39632.2	0.45
$\frac{1}{\lvert k \sim j \rvert} \cdot \mathbb{1}(j \sim i)$	0.30 (0.01)	0.01 (0.02)–0.68 (0.25)	0.03 (0.001)	39631.6	0.44
p_{ji}/n_j	0.28 (0.01)	0.89 (3.13)–31.58 (6.04)	0.03 (0.001)	39617.0	0.45
$p_{ji}(\text{yearly})/n_j$	0.28 (0.01)	0.84 (1.09)–28.68 (5.02)	0.03 (0.001)	39593.5	$*$

The endemic component $\nu_{i,t}$ includes $S=4$ seasonal terms. AIC $=-2\log L+2p$ and max EV denotes the maximum eigenvalue of the estimated matrix $\hat{\mathbf{\Lambda}}$ in (7). The maximum eigenvalues of the model with time-varying weights $(*)$ are shown in Figure 5.

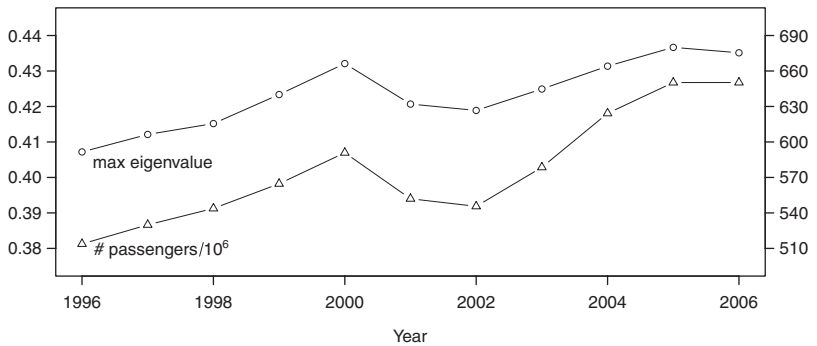

Figure 5. Maximum eigenvalues (circles) of $\hat{\mathbf{\Lambda}}$ for the model with time-varying weights $p_{ji}(\text{yearly})/n_j$ summarized in Table V (left axis). The triangles show the yearly number of passengers per 10^6 (right axis).

unity, the process is ergodic [38]. The largest eigenvalues of $\hat{\mathbf{\Lambda}}$ in the models considered are shown in Table V. Taking spatial variation into account by additional inclusion of $\phi_i \sum w_{ji} y_{j,t-1}$ in the epidemic component leads to an increased maximum eigenvalue compared with the model, which only includes an autoregressive parameter λ. For the model with time-varying weights, the matrix $\hat{\mathbf{\Lambda}}$ changes each year. The respective maximum eigenvalues are shown in Figure 5 together with the yearly number of passengers $\sum p_{ji}(\text{yearly})/10^6$. The development of the number of passengers is mirrored in the curve of the largest eigenvalues.

4. DISCUSSION

In this paper we have proposed a flexible class of statistical models for the analysis of multivariate time series of infectious disease counts. All analyses were done using standard optimization routines where results are readily available in contrast to computer-intensive methods based on Markov chain Monte Carlo (MCMC). A main feature of the model is the decomposition of the disease incidence into an endemic and an epidemic component, which allows to capture occasional

DOI: 10.1002/sim

outbreaks in the data. The motivation of the epidemic component comes from a branching process formulation well known in infectious disease epidemiology. Note that the interpretation of the branching process as an approximation is only appropriate if the generation time, i.e. the time between 'generations' of infectives, equals the observation time at which data are collected (i.e. days, weeks or months). However, simulation studies showed that a Poisson branching process, aggregated to coarser time intervals, can be approximated by a branching process with additional overdispersion [9].

In our example from Section 3.1 the assumption of disease-specific parameters proved to be reasonable. There are much more influenza cases than meningococcal disease cases and therefore the amplitudes of the seasonal patterns differ a lot. Besides, the influenza and meningococcal disease data showed a different amount of overdispersion. The use of weights w_{ji} in the epidemic component may lead to a better accounting for spatial dependence. For instance, the analysis of influenza and pneumonia mortality in nine major regions of the U.S.A. in Section 3.2 showed that the inclusion of air travel information yielded a better model in terms of AIC.

A further model generalization, which we are currently considering, is the introduction of random effects in the endemic or epidemic component (e.g. for incidence levels or for the autoregressive parameters). If the number of units m is large, the number of unit-specific parameters may otherwise get large rather quickly, which might lead to identifiability problems. Estimates of parameters can then be obtained by optimizing the marginal likelihood, which involves integration of the likelihood with respect to the distribution of the random effects. This integral cannot be computed explicitly and must be approximated, e.g. using adaptive Gaussian quadrature or the Laplace approximation for the integrand.

Seasonality is modelled with superposition of sine and cosine terms in the endemic component, which implies that the phase and the amplitude of the seasonal pattern is constant across all considered periods. This might not always be adequate, e.g. if there is a year-by-year variation in the starting point of the seasonal pattern. The formulation is, however, not restricted to this specific modelling of seasonal variation, other choices would also be possible. For example, Fanshawe *et al.* [39] use harmonic regression to model seasonal variations in particulate matter concentrations. To deal with yearly variation of the seasonal pattern the static parameters γ and δ of the sine and cosine terms are replaced with independent random walks.

In this paper a retrospective analysis of time series of counts of infectious diseases was of main interest. The comparison of the different models was based on the model choice criterion AIC. An alternative is to use BIC, in particular if the quality of one-step-ahead predictions is of interest [40]. If the prediction of future observations is a main goal, the validity of models can also be assessed with proper scoring rules that evaluate a model based on the prediction and the actual observed value [41, 42].

Another extension not considered here is the introduction of time-varying parameters, in particular time-varying autoregressive parameters λ_t or ϕ_t as suggested by Held *et al.* [25] where the autoregressive parameter λ is allowed to change over time according to a Bayesian change-point model with unknown number of change-points. The inclusion of a time-varying autoregressive parameter λ_t is especially suited if outbreak detection is the primary focus, which was not the case in this paper. However, such modifications may also be useful in a retrospective analysis if the infectiousness of a disease changes through public health measures such as increasing vaccination coverage or the reduction of infected person's contact rates by prescribed quarantine. If information on such measures is available, a regression approach might be considered, linking λ_t with the known explanatory variables x_t, e.g. through $\lambda_t = \exp(x_t\beta)$. This would have the advantage

DOI: 10.1002/sim

that statistical inference is still possible using ML without the need for a fully Bayesian analysis using MCMC.

APPENDIX A: IMPLEMENTATION DETAILS

The log-likelihood given in Section 2.3 needs to be maximized numerically. We use the quasi-Newton method BFGS to obtain the ML estimates and the corresponding standard errors. To ensure positivity of the dispersion parameters ψ_i and the autoregressive parameters λ_i and ϕ_i, these parameters are optimized on the log-scale, i.e. $\tilde{\psi}_i = \log(\psi_i)$, $\tilde{\lambda}_i = \log(\lambda_i)$ and $\tilde{\phi}_i = \log(\phi_i)$ are used instead. The ML estimates and the corresponding standard errors for the original parameters are obtained using the invariance of the ML estimate and the delta method.

In the following, we will give details for the model without unit-specific parameters, i.e. $\tilde{\boldsymbol{\theta}}_i = (\tilde{\lambda}, \tilde{\phi}, \alpha_i, \gamma_1, \ldots, \gamma_S, \delta_1, \ldots, \delta_S)^{\mathrm{T}}$, $i = 1, \ldots, m$. The treatment of the general case is similar. It is not necessary to supply analytic derivatives of the log-likelihood function in `optim` to obtain the ML estimates. However, as the score function

$$s(\tilde{\boldsymbol{\theta}}) = \frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} l(\tilde{\boldsymbol{\theta}}) = \sum_{i,t} \frac{\partial}{\partial \tilde{\boldsymbol{\theta}}} l_{i,t}(\tilde{\boldsymbol{\theta}}_i, \psi)$$

can be derived analytically for our model, we use this information since it speeds up the computation of the ML estimates considerably.

Denote an element of the parameter vector $\tilde{\boldsymbol{\theta}}$ as $\tilde{\theta}_k$; the score function contribution of observation $y_{i,t}$ is then given as

$$\frac{\partial}{\partial \tilde{\theta}_k} l_{i,t}(\tilde{\boldsymbol{\theta}}_i, \tilde{\psi}) = -\frac{\exp(-\tilde{\psi})}{\exp(-\tilde{\psi}) + \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)} \cdot \frac{\partial}{\partial \tilde{\theta}_k} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) + \frac{y}{\mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)} \cdot \frac{\partial}{\partial \tilde{\theta}_k} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)$$

$$- \frac{y_{i,t}}{\exp(-\tilde{\psi}) + \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)} \cdot \frac{\partial}{\partial \tilde{\theta}_k} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)$$

$$\frac{\partial}{\partial \tilde{\psi}} l_{i,t}(\tilde{\boldsymbol{\theta}}_i, \tilde{\psi}) = (-\Psi(y_{i,t} + \exp(-\tilde{\psi})) + \Psi(\exp(-\tilde{\psi})) - \log(\exp(-\tilde{\psi})) - 1$$

$$+ \log(\exp(-\tilde{\psi}) + \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)) + \frac{\exp(-\tilde{\psi}) + y_{i,t}}{\exp(-\tilde{\psi}) + \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i)}) \cdot \exp(-\tilde{\psi})$$

where $\Psi(z) = \mathrm{d}\log(\Gamma(z))/\mathrm{d}z$ is the digamma function [26, p. 258] and

$$\frac{\partial}{\partial \tilde{\lambda}} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) = \exp(\tilde{\lambda}) \cdot y_{i,t-1}$$

$$\frac{\partial}{\partial \tilde{\phi}} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) = \exp(\tilde{\phi}) \cdot \sum_{j \neq i} w_{ji} y_{j,t-l}$$

$$\frac{\partial}{\partial \alpha} \mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) = v_{i,t}$$

DOI: 10.1002/sim

$$\frac{\partial}{\partial \gamma_s}\mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) = v_{i,t}\cdot\sin(\omega_s t)$$

$$\frac{\partial}{\partial \delta_s}\mu_{i,t}(\tilde{\boldsymbol{\theta}}_i) = v_{i,t}\cdot\cos(\omega_s t)$$

The observed Fisher information matrix $F(\tilde{\boldsymbol{\theta}}) = -\partial s(\tilde{\boldsymbol{\theta}})/\partial\tilde{\boldsymbol{\theta}}$ can also be derived analytically and used to obtain standard errors of the ML estimates.

Newton–Raphson-type methods for optimization require that the initial values are sufficiently close to the solution to guarantee convergence. Far away from the solution, the algorithm can diverge. In addition, it is possible that the optimization algorithm converges to points which are not global optima. Poor convergence or convergence to such local optima occurs much more frequently in multivariate problems involving many parameters than in univariate problems [43]. It is therefore recommended to try multiple starting values to find the global maximum. Even so, results are nearly instantly available.

ACKNOWLEDGEMENTS

We thank the referees for helpful comments and suggestions. Financial support by the Swiss National Science Foundation (SNF) is gratefully acknowledged.

REFERENCES

1. Daley DJ, Gani J. *Epidemic Modelling*: *An Introduction*. Cambridge University Press: Cambridge, 1999.
2. Andersson H, Britton T. *Stochastic Epidemic Models and their Statistical Analysis*. Lecture Notes in Statistics, vol. 151. Springer: New York, 2000.
3. Griffiths DA. Multivariate birth-and-death processes as approximations to epidemic processes. *Journal of Applied Probability* 1973; **10**(1):15–26.
4. Cox DR, Donnelly CA, Bourne FJ, Gettinby G, McInerney JP, Morrison WI, Woodroffe R. Simple model for tuberculosis in cattle and badgers. *Proceedings of the National Academy of Sciences of the United States of America* 2005; **102**(49):17588–17593.
5. Finkenstädt BF, Bjørnstad ON, Grenfell BT. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics* 2002; **3**(4):493–510.
6. Finkenstädt BF, Grenfell BT. Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society*, *Series C*: *Applied Statistics* 2000; **49**(2):187–205.
7. Zeger SL, Qaqish B. Markov regression models for time series: a quasi-likelihood approach. *Biometrics* 1988; **44**(4):1019–1031.
8. Knorr-Held L, Richardson S. A hierarchical model for space–time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society*, *Series C*: *Applied Statistics* 2003; **52**(2):169–183.
9. Held L, Höhle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* 2005; **5**:187–199.
10. Hament JM, Kimpen JL, Fleer A, Wolfs TF. Respiratory viral infection predisposing for bacterial disease: a concise review. *FEMS Immunology and Medical Microbiology* 1999; **26**(3–4):189–195.
11. Cartwright KA, Jones DM, Smith AJ, Stuart JM, Kaczmarski EB, Palmer SR. Influenza A and meningococcal disease. *The Lancet* 1991; **338**(8766):554–557.
12. Hubert B, Watier L, Garnerin P, Richardson S. Meningococcal disease and influenza-like syndrome: a new approach to an old question. *The Journal of Infectious Diseases* 1992; **166**(3):542–545.
13. Jensen ES, Lundbye-Christensen S, Samuelsson S, Sørensen HT, Schønheyder HC. A 20-year ecological study of the temporal association between influenza and meningococcal disease. *European Journal of Epidemiology* 2004; **19**(2):181–187.

DOI: 10.1002/sim

14. Jansen AGSC, Sanders EAM, Van Der Ende A, Van Loon AM, Hoes AW, Hak E. Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity? *Epidemiology and Infection* 2008; DOI: 10.1017/S0950268807000271.
15. Robert Koch Institute. SurvStat. Available from: http://www3.rki.de/SurvStat. Accessed July 2007.
16. Farrington CP, Kanaan MN, Gay NJ. Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Journal of the Royal Statistical Society*, *Series C*: *Applied Statistics* 2001; **50**:251–283.
17. De Angelis D, Gilks WR, Day NE. Bayesian projection of the acquired immune deficiency syndrome epidemic. *Journal of the Royal Statistical Society*, *Series C*: *Applied Statistics* 1998; **47**:449–481.
18. Brownstein JS, Wolfe CJ, Mandl KD. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine* 2006; **3**(10):e401.
19. Grais RF, Ellis JH, Glass GE. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology* 2003; **18**(11):1065–1072.
20. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* 2004; **101**(42):15124–15129.
21. Colizza V, Barrat A, Barthélemy M, Vespignani A. The modeling of global epidemics: stochastic dynamics and predictability. *Bulletin of Mathematical Biology* 2006; **68**(8):1893–1921.
22. U.S. Department of Transportation, Bureau of Transportation Statistics. Database: Air Carrier Statistics (Form 41 Traffic): T-100 Domestic Segment (All Carriers). Available from: http://www.transtats.bts.gov. Accessed June 2007.
23. Serfling R. Methods for current statistical analysis of excess pneumonia–influenza deaths. *Public Health Reports* 1963; **78**:494–506.
24. Le Strat Y, Carrat F. Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine* 1999; **18**(24):3463–3478.
25. Held L, Hofmann M, Höhle M, Schmid V. A two-component model for counts of infectious diseases. *Biostatistics* 2006; **7**(3):422–437.
26. Abramowitz M, Stegun IA. *Handbook of Mathematical Functions with Formulas*, *Graphs*, *and Mathematical Tables*. Dover: New York, 1964.
27. Thurston SW, Wand MP, Wiencke JK. Negative binomial additive models. *Biometrics* 2000; **56**(1):139–144.
28. Höhle M. Surveillance: an R package for the surveillance of infectious diseases (2007). *Computational Statistics* 2007; **22**(4):571–582.
29. Brundage JF. Interactions between influenza and bacterial respiratory pathogens: implications for pandemic preparedness. *The Lancet Infectious Diseases* 2006; **6**(5):303–312.
30. Schrauder A, Claus H, Elias J, Vogel U, Haas W, Hellenbrand W. Capture–recapture analysis to estimate the incidence of invasive meningococcal disease in Germany, 2003. *Epidemiology and Infection* 2007; **135**(4):657–664.
31. Svetliza CF, Paula GA. Diagnostics in nonlinear negative binomial models. *Communications in Statistics. Theory and Methods* 2003; **32**(6):1227–1250.
32. Diggle PJ. *Time Series. A Biostatistical Introduction*. Oxford University Press: Oxford, 1990.
33. Diggle PJ, Heagerty P, Liang KY, Zeger S. *The Analysis of Longitudinal Data* (*Oxford Statistical Science*). Oxford University Press: Oxford, 2002.
34. Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. *Journal of the American Statistical Association* 2003; **98**(461):214–223.
35. Davis RA, Dunsmuir WTM, Streett SB. Observation-driven models for Poisson counts. *Biometrika* 2003; **90**(4): 777–790.
36. Centers for Disease Control and Prevention. MMWR Table III (Mortality). Available from: http://www.cdc.gov/EPO/DPHSI/121hist.htm. Accessed December 2006.
37. US Census Bureau, Population Division. Population, population change and estimated components of population change: April 1, 2000 to July 1, 2006 (NST_EST2006_ALLDATA). Available from: http://www.census.gov/popest/datasets.html. Accessed July 2007.
38. Mode CJ. *Multitype Branching Processes—Theory and Applications*. American Elsevier Publishing Company: New York, 1971.
39. Fanshawe T, Diggle P, Rushton S, Sanderson R, Lurz P, Glinianaia S, Pearce M, Parker L, Charlton M, Pless-Mulloli T. Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics* 2008; **19**(6):549–566.

DOI: 10.1002/sim

40. Dawid AP. Statistical theory. The prequential approach. *Journal of the Royal Statistical Society*, *Series A*: *General* 1984; **147**(2):278–292.
41. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 2007; **102**(477):359–378.
42. Czado C, Gneiting T, Held L. Predictive model assessment for count data. *Technical Report 518*, Department of Statistics, University of Washington, 2007.
43. Thisted RA. *Elements of Statistical Computing. Numerical Computation*. Chapman & Hall: New York, 1988.

DOI: 10.1002/sim