Monitoring Poisson time series using multi-process models

Engebjerg, Malene Dahl Skov; Lundbye-Christensen, Søren; Kjær, Birgitte B.; Schønheyder, Henrik Carl

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

AALBORG UNIVERSITY

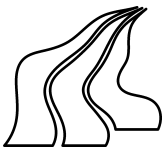
Monitoring Poisson time series using multi-process models

by

Malene C. Engebjerg, Søren Lundbye-Christensen,
Birgitte B. Kjær and Henrik C. Schønheyder

R-2006-07 March 2006

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: http://www.math.aau.dk



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Monitoring Poisson Time Series using Multi-Process Models

Malene C. Engebjerg

Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark

Søren Lundbye-Christensen

Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark

Birgitte B. Kjær

Department of Bacteriology, Mycoplogy and Parasitology, Statens Serum Institut, Copenhagen, Denmark

Henrik C. Schønheyder

Department of Clinical Microbiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

Abstract. Surveillance of infectious diseases based on routinely collected public health data is important for at least three reasons: The early detection of an epidemic may facilitate prompt interventions and the seasonal variations and long term trend may be of general epidemiological interest. Furthermore aspects of health resource management may also be addressed. In this paper we center on the detection of outbreaks of infectious diseases. This is achieved by a multi-process Poisson state space model taking autocorrelation and overdispersion into account, which has been applied to a data set concerning Mycoplasma pneumoniae infections.

Keywords: Infectious disease, Outbreak, State space models, Surveillance, Warning system.

1. Introduction

Monitoring of routinely collected incidences of infectious diseases is of great importance in public health service. There may be at least three reasons for this. First of all early detection of the onset of an epidemic may provide the public health authorities with the information needed to make appropriate interventions. Secondly changes in seasonality or general trend of a disease may be of epidemiological interest and finally aspects of resource management and quality control should not be underestimated. However, the number of health-related information systems and the sheer amount of data available are beyond the limits of manual surveillance and hence automated monitoring procedures are called for.

Farrington et al. (1996) describe an algorithm for the detection of outbreaks of infectious disease now being used at Statens Serum Institut, Denmark, in the monitoring of the gastointestinal pathogens Salmonella, Campylobacter, Yersinia, Shigella and E. coli, see SSI (2005). Farrington et al. (1996) point out that timeliness, sensitivity and specificity are the main objectives of such an algorithm. This is accomplished through a log-linear

Address for correspondence: Malene C. Engebjerg, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark
E-mail: maleneds@math.aau.dk

regression model, adjusted for overdispersion, seasonality, secular trends and past outbreaks. The model is then used to calculate a threshold-value above which any observed counts are flagged and further investigated. The threshold-value is calculated on the basis of corresponding weeks from previous years thereby incorporating the seasonal variation in the data. However, if some of these weeks correspond to past outbreaks the threshold-value will be too high and hence the sensitivity declines. Consequently a weight function is defined giving low weight to weeks with unusually high counts and in that way the threshold-value is corrected for previous outbreaks.

A number of medical time series have been analyzed by Gordon and Smith (1990). The requirement was the detection of events of clinical or biological importance. This was achieved using a multi-process Gaussian model, see Section 2.3, capable of handling missing values. In particular an application for monitoring renal transplants must be noted (a more detailed description can be found in Smith and West (1983)).

Whittaker and Frühwirth-Schnatter (1994) use a multi-process Gaussian model to detect the possible time point at which bacteriological growth takes place in the monitoring of feedstuff. Whether one should stop sampling because bacteriological infection is highly likely is determined by a decision rule which is defined in terms of the costs for false negatives and false positives.

Strat and Carrat (1999) illustrate how hidden Markov models may be used in monitoring surveillance data. The approach actually mimics the multi-process models presented in the current paper, Section 2.3 and 3, in that the observed series depends on an underlying state. Typically there are two states corresponding to a non-epidemic and an epidemic situation. The states are assumed to follow a homogeneous Markov chain. There are, however, substantial differences between the two approaches. First of all Strat and Carrat (1999) adopt a non-Bayesian and a non-dynamic approach. The latter implies that the trend and seasonality are stationary quantities and hence do not vary over time. Furthermore the method does not naturally incorporate on-line updating which is desirable when it comes to detection of outbreaks.

Cooper and Lipsitch (2004) go a step further and propose a structured hidden Markov model in the analysis of hospital infection data still in a non-Bayesian and non-dynamic setup. The motive for this is that hospital infections are often only partially observed. This is overcome be letting the states correspond to the true number of patients who has harbored the infection as opposed to the two (epidemic and non-epidemic states) proposed by Strat and Carrat (1999). The distribution of the observed series now depend on the unobserved state by assuming that the mean is proportional to the true number of infected patients. In this way autocorrelation and overdispersion in the observed series is taken care of. The term *structured* stems from the fact that a mechanistic understanding of the biological and epidemiological nature of the data is incorporated into the model. This is achieved by defining the transition probabilities between the states in accordance with the stochastic susceptible-infectious-susceptible epidemic model. This has the clear advantage that biological and epidemiological parameters may be estimated. This may prove to be valuable in the understanding of the aetiology of an infection.

Finkenstädt and Grenfell (2000) propose a time series susceptible-infected-recovered epidemic model for the analysis of measles data from England and Wales from 1944 to 1964. In that way they also incorporate a mechanistic approach by defining a recursive stochastic relationship between the number of infected and susceptible individuals. Due to under-reporting the number of infected individuals are only partially observed, whereas the number of susceptible individuals are completely unknown. The model is fitted by

reconstructing the number of susceptible individuals using locally linear regression. As opposed to Cooper and Lipsitch (2004) this approach is dynamic and better suited for population modeling.

A much less complicated surveillance system based on weighted moving averages is presented by Dessau and Steenberg (1993). It is intended for on-line surveillance and implemented in a laboratory information system.

The objective of this paper has been to present a model which concisely incorporates many of the advantages of the above mentioned methods. This has been achieved by a multi-process Poisson model, where autocorrelation and overdispersion is handled via a latent process. The modeling of change-points is facilitated by the multi-process approach. Furthermore the approach is Bayesian, which allows for expert prior information to be directly incorporated into the model. Finally a rigid description of the base activity is avoided by letting the regression parameters vary randomly over time. The model is conceptually simple in nature and the setup is based on sequential updating equations which conform nicely to the requirement for on-line surveillance. In this paper, however, we have not adopted a mechanistic approach but only focused on surveillance and warning. We apply this model to Danish nationwide laboratory data concerning Mycoplasma pneumoniae infections.

1.1. Layout

In Section 2 we give a brief review of the basic state space models. In particular we will define the *dynamic linear model (DLM)* (also denoted Gaussian state space model), where the observed series is assumed to be conditionally Gaussian distributed given a latent process. We then proceed a step further by outlining the case, where the Gaussian assumption is relaxed obtaining a *dynamic generalized linear model (DGLM)*. Finally a *multi-process Gaussian model* is presented, which extends the dynamic linear model by allowing any of a finite number of DLMs to describe the observation at any given time point.

Section 3 combines the dynamic generalized linear model with the multi-process Gaussian model to obtain a *multi-process Poisson model*, where one of the states is particularly designed to capture outliers. In this section we will give a detailed account of the updating procedures used in the multi-process Poisson model. In the implementation of the multi-process Poisson model some computational issues arose, which will be addressed in Section 4. Section 5 contains an application to Mycoplasma pneumoniae infections and we end with a discussion in Section 6.

2. General State Space Models

Consider the time series

$$y_1, y_2, \ldots, y_t, \ldots$$

observed at equally spaced time intervals, where y_t is assumed to be a realization of Y_t for all t. The assumption of equispaced observations can easily be relaxed. Generally the observations are thought to be autocorrelated as opposed to an independent and identically distributed sample. In the present context the autocorrelation is partly due to the communicable nature of infectious diseases.

State space models provide a very general setup in which this autocorrelation is modeled via a latent process, $\boldsymbol{\theta}_t$. The idea is that the observed series is considered conditional

independent given $\boldsymbol{\theta}_t$. Apart from the fact that a time series usually is autocorrelated one often finds that a rigid description of the underlying mean is unsatisfactory. This is accomplished in the state space setup by allowing the effect of the explanatory variables to evolve randomly over time.

An extension of the state space model as outlined above is the multi-process state space model, which is achieved by combining a finite collection of state space models. Hence we do not expect any single state space model to describe the data sufficiently throughout the study period. The multi-process model was first proposed by Harrison and Stevens (1971) and a more general account is given by Harrison and Stevens (1976).

2.1. Dynamic Linear Models

The model setup is as follows:

$$Y_t = F_t^T \boldsymbol{\theta_t} + \nu_t \qquad \nu_t \sim N[0, V_t] \tag{1}$$

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N[0, W_t] \tag{2}$$

with prior information given by

$$(\boldsymbol{\theta}_0 \,|\, D_0) \sim N[\boldsymbol{m}_0, C_0]. \tag{3}$$

Here F_t is a $p \times 1$ regression vector, $\boldsymbol{\theta_t}$ is a $p \times 1$ parameter vector and G_t is a $p \times p$ evolution matrix. The error series $\{\nu_t, \omega_t; t = 1, 2, \dots\}$ is assumed to be mutually independent. We note that $V_t, W_t, F_t, G_t, \boldsymbol{m}_0$ and C_0 are specified by the modeler.

Furthermore D_t denotes any historical information at time t relevant for the system including $V_t, W_t, F_t, G_t, \boldsymbol{m}_0$ and C_0. Furthermore we will assume that any future information set D_t will be closed to external information, i.e.

$$D_t = \{y_t, D_{t-1}\} \quad \text{for} \quad t = 1, 2, \dots.$$

An account of the recursive updating scheme used for this univariate dynamic linear model can be found in West and Harrison (1999).

2.2. Dynamic Generalized Linear Models

We assume in the dynamic linear model that $p(Y_t \,|\, \boldsymbol{\theta}_t, V_t)$ is a Gaussian density. In the dynamic generalized linear model this density is replaced with a general density from the exponential family, i.e.

$$p(y_t \,|\, \eta_t, V_t) = \exp(V_t^{-1}(y_t \eta_t - a(\eta_t))) b(y_t, V_t).$$

We then propose the following model:

$$g(\mu_t) = \lambda_t = F_t^T \boldsymbol{\theta}_t$$

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim [0, W_t]$$

with prior information $(\boldsymbol{\theta}_0 | D_0) \sim [\boldsymbol{m}_0, C_0]$, where $[\cdot, \cdot]$ denotes a distribution which is only partially specified through the mean and variance. Again the error series $\{\omega_t; t = 1, 2, \dots\}$ is assumed to be mutually independent.

Most often the canonical link function g will be used, that is to say the function g satisfying $g(\mu_t) = \eta_t$. In the Poisson case $g(\cdot) = \log(\cdot)$. See West and Harrison (1999) for a treatment of the dynamic generalized linear model.

2.3. Multi-Process Gaussian Models

The multi-process Gaussian models allow for sudden changes in pattern, not described by the random variation in the latent process. This is accomplished by augmenting the dynamic linear model with a discrete process, M_t indicating which one in a range of possible models applies at a given time. This process could e.g. distinguish between an epidemic and a non-epidemic situation. The description of aberrant states is usually done by specifying larger evolutional or observational variances.

Furthermore we assume that model $M_t(i_t)$ pertains at time t with probability $\pi(i_t)$ irrespective of the past, i.e.

$$\mathrm{P}(M_t \mid M_{t-1}, \ldots, M_1, D_{t-1}) = \mathrm{P}(M_t \mid D_0) = \pi(i_t).$$

This is also known as fixed model selection probabilities. In some applications a first or higher order Markov model for M_t might be appropriate.

3. Multi-Process Poisson Models

In applications to count-data – e.g. the number of disease occurrences in a given time period – the normal assumption will typically be inappropriate and therefore a multi-process Poisson model is presented in the current section. Here we will combine the methods given in Section 2.2 and 2.3 to specify a multi-process model in the case where the observations are assumed to be Poisson distributed. The combination of the dynamic generalized linear model and the multi-process Gaussian model is straightforward except for the specification of the outlier model. This is due to the fact that the dispersion parameter in the Poisson distribution is 1. On the other hand it is essential in public health surveillance to be able to distinguish an outlier from a permanent shift in the latent process. This stems from the fact that the accumulation of sporadic cases is immaterial in the detection of an epidemic.

In an unpublished research report West (1986) presents another approach to the multi-process model which applies much more generally in that the observations are merely assumed to follow an exponential family sampling distribution. On the other hand the outlier model specified by West differs considerably from the outlier model offered in this paper. The main difference is that the sampling model is state independent in West (1986), whereas we propose a state dependent sampling model specially designed for the Poisson case and in that way this is more reminiscent of the multi-process Gaussian model.

Yet another approach to the multi-process Poisson model has been made by Bolstad (1995). Here the modeling is restricted to the case, where the logarithm of the intensity is described by a straight line and the states correspond to (1) steady state, (2) level change and (3) outlier. In that way the handling of the different states is closely tied together with the actual model proposed.

The model setup in this paper is as follows.

For $i_t = 1, 2, \ldots, N$ (N being the number of different states) we define

$$Y_t \mid \mu_t, M_t = i_t \sim \mathrm{Pois}(\mu_t \Delta_{i_t}) \tag{4}$$

$$\log(\mu_t) = \eta_t = F_t^T \boldsymbol{\theta_t} \tag{5}$$

$$\boldsymbol{\theta_t} = G_t \boldsymbol{\theta_{t-1}} + \boldsymbol{\omega_t}, \qquad \boldsymbol{\omega_t} \mid M_t = i_t \sim [0, W_t(i_t)], \tag{6}$$

with fixed model selection probabilities π. The error series $\{\omega_t; t = 1, 2, \ldots\}$ is assumed conditional independent given all model states.

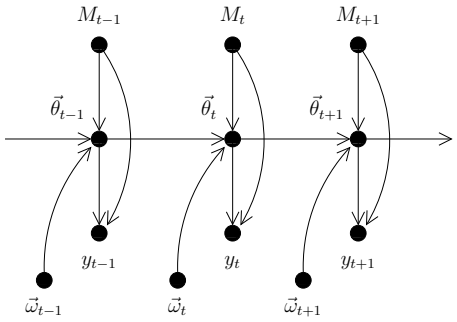


Figure 1. Directed Acyclic Graph (DAG) from which conditional independence relations may be identified.

One may prove that the joint distribution as specified above admits a recursive factorization according to the graph shown in Figure 1. Any conditional independence relations stated in the following can readily be identified from this graph, see Lauritzen (1996) or Edwards (2000) for an account on graphical models. The parameter Δ_{i_t} is the main ingredient in the outlier model. For any non-outlier state $\Delta_{i_t} = 1$, a large outlier is specified by choosing $\Delta_{i_t} > 1$ and similarly a small outlier by setting $\Delta_{i_t} < 1$. This implies that the forecaster needs to define two or more Δ_{i_t}-values if he wants to be able to predict both small and large outliers. This may seem as a drawback. There is, however, at least two reasons which justify this outlier model. The outlier model is transparent and the size of Δ_{i_t} corresponds to the amount by which a potential outlier is expected to be above or below the mean value. The Poisson model is primarily necessary when the observed counts are expected to be relatively small. Whenever this is the case small outliers will be indistinguishable from the remaining observations and there will be no need for Δ_{i_t}-values less than 1. In all other cases the multi-process Gaussian model will usually provide a satisfactory approximation.

The variances $W_t(i_t)$ are once again used to specify the steady state model as well as any change model the forecaster finds appropriate. A sudden change in e.g. growth rate is modeled by an appropriate elevation of the evolutional variance corresponding to the slope component.

As in the multi-process Gaussian model the prior for $\boldsymbol{\theta}_t$ is only partially specified through its mean and variance. However, we will later force a conjugate prior upon μ_t given information on the last $k+1$ models. This will basically superimpose distributional bindings upon all the above parameters although an exact distribution of these will be far too complicated to take into account.

The updating scheme – also known as the multi-process Kalman filter – has been decomposed into 8 steps all of which will be described in the sequel.

3.1. Sequential Updating Equations

For notational convenience we will in the following abbreviate $M_t = i_t$ with M_t, $(M_t, \ldots, M_{t-j})$ with M_{t-j}^t and similarly $(i_t, \ldots, i_{t-j})$ with i_{t-j}^t.

Input to the updating scheme at time t
Beyond the regression vector F_t, the evolution matrix G_t, the covariance matrices $\{W_t(i_t)\}_{i_t=1}^N$ and the model selection probabilities $\{\pi(i_t)\}_{i_t=1}^N$ we assume that

$$(\boldsymbol{\theta}_{t-1} \mid M_{t-k}^{t-1}, D_{t-1}) \sim \left[\boldsymbol{m}_{t-1}(i_{t-k}^{t-1}), C_{t-1}(i_{t-k}^{t-1})\right]$$

and that the joint probability $p(M_{t-k}^{t-1} \mid D_{t-1})$ is known. Note here that k denotes the lag time.

Initialization
If $t = 1$ we let

$$(\boldsymbol{\theta}_0 \mid M_{1-k}^0, D_0) \equiv (\boldsymbol{\theta}_0 \mid M_0, D_0) \sim [\boldsymbol{m}_0, C_0]$$

and $p(M_{1-k}^0 \mid D_0)$ is left undefined.

The iterative updating scheme then proceeds as follows.

Step 1 [Conditional prior for $\boldsymbol{\theta}_t$ and η_t]
Using (5) and (6) the moments

$$\begin{aligned}(\boldsymbol{\theta}_t \mid M_{t-k}^t, D_{t-1}) &\sim [\boldsymbol{a}_t(i_{t-k}^{t-1}), R_t(i_{t-k}^t)],\\ (\eta_t \mid M_{t-k}^t, D_{t-1}) &\sim [f_t(i_{t-k}^{t-1}), q_t(i_{t-k}^t)]\end{aligned} \tag{7}$$

and

$$\mathbb{C}\mathrm{ov}\,(\eta_t, \boldsymbol{\theta}_t \mid M_t(i_{t-k}^t), D_{t-1}) = F_t^T R_t(i_{t-k}^t).$$

can easily be found.

Step 2 [Conjugate conditional prior for μ_t]
To get a conjugate updating scheme we premise $(\mu_t \mid M_{t-k}^t, D_{t-1})$ to be Gamma distributed with, say, form parameter $r_t(i_{t-k}^t)$ and scale parameter $s_t(i_{t-k}^t)$. This implies that

$$p(\eta_t \mid M_{t-k}^t, D_{t-1}) \propto \exp(r_t(i_{t-k}^t)\eta_t - s_t(i_{t-k}^t)\exp(\eta_t)).$$

One must now choose the parameters $r_t(i_{t-k}^t)$ and $s_t(i_{t-k}^t)$ such that the prior moments for η_t are as in (7). Making a first order Taylor expansion of $\mu_t(i_{t-k}^t)$ around $\exp(f_t(i_{t-k}^{t-1}))$ and using the mean and variance of $\mu_t(i_{t-k}^t)$ we get

$$\exp(f_t(i_{t-k}^{t-1})) \simeq \frac{r_t(i_{t-k}^t)}{s_t(i_{t-k}^t)} \quad \text{and} \quad q_t(i_{t-k}^t)\exp(2f_t(i_{t-k}^{t-1})) = \frac{r_t(i_{t-k}^t)}{s_t(i_{t-k}^t)^2},$$

which lead us to choose $r_t(i_{t-k}^t)$ and $s_t(i_{t-k}^t)$, for each combination of states i_{t-k}^t, in the following way

$$\begin{aligned}r_t(i_{t-k}^t) &= (q_t(i_{t-k}^t))^{-1}\\ s_t(i_{t-k}^t) &= (\exp(f_t(i_{t-k}^{t-1})) \cdot q_t(i_{t-k}^t))^{-1}.\end{aligned}$$

Step 3 [Conditional predictive distribution for y_t]
The predictive distribution for y_t is found by marginalization in the joint distribution for $(y_t, \mu_t \,|\, M_{t-k}^t, D_{t-1})$:

$$p(y_t \,|\, M_{t-k}^t, D_{t-1}) = \int p(y_t|\mu_t, M_{t-k}^t, D_{t-1}) p(\mu_t \,|\, M_{t-k}^t, D_{t-1}) \mathrm{d}\mu_t.$$

The former density on the right only depends on (μ_t, M_t) and is Poisson with parameter $\mu_t \Delta_{i_t}$. Since the latter density is conjugate to the Poisson distribution the integrand is seen to be proportional to a $\Gamma(y_t + r_t(i_{t-k}^t), \Delta_{i_t} + s_t(i_{t-k}^t))$-density and hence the normalization constant can easily be found. Therefore the predictive density is

$$p(y_t \,|\, M_{t-k}^t, D_{t-1}) = \Delta_{i_t}^{y_t} \cdot \frac{1}{y_t!} \cdot \frac{\Gamma(y_t + r_t(i_{t-k}^t))}{\Gamma(r_t(i_{t-k}^t))} \cdot \frac{(s_t(i_{t-k}^t))^{r_t(i_{t-k}^t)}}{(\Delta_{i_t} + s_t(i_{t-k}^t))^{y_t + r_t(i_{t-k}^t)}}.$$

Step 4 [Conjugate conditional posterior for μ_t]
The conditional posterior for μ_t is given by $(\mu_t \,|\, M_{t-k}^t, D_t) \sim \Gamma(y_t + r_t(i_{t-k}^t), \Delta_{i_t} + s_t(i_{t-k}^t))$ and therefore the posterior moments for η_t are explicitly given in accordance with this relation although difficult to derive in closed form. However, Taylor expanding $\log(\mu_t)$ around its posterior mean yields

$$(\eta_t \,|\, M_{t-k}^t, D_t) \sim [f_t^*(i_{t-k}^t), q_t^*(i_{t-k}^t)],$$

where

$$f_t^*(i_{t-k}^t) \simeq \log\Big(\frac{y_t + r_t(i_{t-k}^t)}{\Delta_{i_t} + s_t(i_{t-k}^t)}\Big) \qquad \text{and} \qquad q_t^*(i_{t-k}^t) \simeq (y_t + r_t(i_{t-k}^t))^{-1}.$$

Step 5 [Conditional prior for $(\boldsymbol{\theta}_t \,|\, \eta_t)$]
Recall from step 1 that

$$\begin{pmatrix} \boldsymbol{\theta}_t \\ \eta_t \end{pmatrix} \,\Big|\, M_{t-k}^t, D_{t-1} \Big) \sim \left[\begin{pmatrix} \boldsymbol{a}_t(i_{t-k}^{t-1}) \\ f_t(i_{t-k}^{t-1}) \end{pmatrix}, \begin{pmatrix} R_t(i_{t-k}^t) & R_t(i_{t-k}^t) F_t \\ F_t^T R_t(i_{t-k}^t) & q_t(i_{t-k}^t) \end{pmatrix} \right].$$

As this prior fails to be specified completely the moments for $(\boldsymbol{\theta}_t \,|\, \eta_t, M_{t-k}^t, D_{t-1})$ are unknown. However, using linear Bayes' estimation, see West and Harrison (1999)[pp. 122], we may approximate the mean and variance by

$$(\boldsymbol{\theta}_t \,|\, \eta_t, M_{t-k}^t, D_{t-1}) \sim [\boldsymbol{a}_t^*(i_{t-k}^t), R_t^*(i_{t-k}^t)],$$

where

$$\boldsymbol{a}_t^*(i_{t-k}^t) \simeq \boldsymbol{a}_t(i_{t-k}^{t-1}) + R_t(i_{t-k}^t) F_t (\eta_t - f_t(i_{t-k}^{t-1}))/q_t(i_{t-k}^t)$$

and

$$R_t^*(i_{t-k}^t) \simeq R_t(i_{t-k}^t) - R_t(i_{t-k}^t) F_t F_t^T R_t(i_{t-k}^t)/q_t(i_{t-k}^t).$$

Note that this corresponds to what we would have found had the above prior been Gaussian.

Step 6 [Posterior model probabilities for M_{t-k}^t]
For notational convenience we let

$$p_t(i_{t-j}^t) = p(M_{t-j}^t \mid D_t), \qquad j = 1, 2, \ldots, k.$$

To update the posterior probability $p_t(i_{t-k}^t)$ note that

$$\begin{aligned} p_t(i_{t-k}^t) &\propto p(y_t \mid M_{t-k}^t, D_{t-1}) p(M_{t-k}^t \mid D_{t-1}) \\ &= p(y_t \mid M_{t-k}^t, D_{t-1}) p(M_t \mid M_{t-k}^{t-1}, D_{t-1}) p(M_{t-k}^{t-1} \mid D_{t-1}). \end{aligned}$$

Since M_t is independent of the past the above reduces to

$$p_t(i_{t-k}^t) \propto p(y_t \mid M_{t-k}^t, D_{t-1}) p_{t-1}(i_{t-k}^{t-1}) \pi(i_t). \tag{8}$$

The first term on the right hand side was found in step 3 and the last term is specified by the forecaster. At last the probability $p_{t-1}(i_{t-k}^{t-1})$ was given as input to the algorithm.

Step 7 [Posterior model probabilities for M_{t-k+1}^t]
At time $t+1$ the updating scheme assumes the conditional probability $p_t(i_{t-k+1}^t)$ to be known which easily follows from step 6:

$$p_t(i_{t-k+1}^t) = \sum_{i_{t-k}=1}^{N} p_t(i_{t-k}^t).$$

Step 8 [Conditional posterior for $\boldsymbol{\theta}_t$]
Finally the posterior mean and variance for $\boldsymbol{\theta}_t$ must be provided for the updating at time $t+1$. Suppose that $(\boldsymbol{\theta}_t \mid M_{t-k+1}^t, D_t)$ has a density. Then

$$p(\boldsymbol{\theta}_t \mid M_{t-k+1}^t, D_t) = \sum_{i_{t-k}=1}^{N} p(\boldsymbol{\theta}_t \mid M_{t-k}^t, D_t) \frac{p_t(i_{t-k}^t)}{p_t(i_{t-k+1}^t)},$$

where the weights $p_t(i_{t-k}^t)/p_t(i_{t-k+1}^t)$ were found in step 6 and 7. Inserting the linear Bayes' estimate from step 5 yields

$$(\boldsymbol{\theta}_t \mid M_{t-k+1}^t, D_t) \sim [\boldsymbol{m}_t(i_{t-k+1}^t), C_t(i_{t-k+1}^t)],$$

where

$$\boldsymbol{m}_t(i_{t-k+1}^t) = \sum_{i_{t-k}=1}^{N} \frac{p_t(i_{t-k}^t)}{p_t(i_{t-k+1}^t)} \boldsymbol{a}_t^*(i_{t-k}^t)$$

and

$$\begin{aligned} C_t(i_{t-k+1}^t) &= \sum_{i_{t-k}=1}^{N} \frac{p_t(i_{t-k}^t)}{p_t(i_{t-k+1}^t)} \int (\boldsymbol{\theta}_t - \boldsymbol{m}_t(i_{t-k+1}^t))^2 p(\boldsymbol{\theta}_t \mid M_{t-k}^t, D_t) \mathrm{d}\boldsymbol{\theta}_t \\ &= \sum_{i_{t-k}=1}^{N} \frac{p_t(i_{t-k}^t)}{p_t(i_{t-k+1}^t)} \cdot \Big(R_t^*(i_{t-k}^t) + (\boldsymbol{a}_t^*(i_{t-k}^t) - \boldsymbol{m}_t(i_{t-k+1}^t))^2 \Big). \end{aligned}$$

Here $\boldsymbol{x}^2 = \boldsymbol{x}\boldsymbol{x}^T$. Since the distribution of $(\boldsymbol{\theta}_t \,|\, M_{t-k+1}^t, D_t)$ is only partially specified via the first and second order moments there is no issue of collapsing densities here as opposed to the updating scheme for the multi-process Gaussian model.

3.2. Output at time t

The previous section presented the recursive updating equations used to run the multi-process Kalman filter. However, other outputs might be of interest in an application of the multi-process Poisson model some of which will be given below.

j-step back model probabilities

The j-step back model probabilities are

$$p(M_{t-j} \,|\, D_t) = \sum_{i_{t-k}=1}^{N} \cdots \sum_{i_{t-j-1}=1}^{N} \sum_{i_{t-j+1}=1}^{N} \cdots \sum_{i_t=1}^{N} p_t(i_{t-k}^t)$$

for $i_{t-j} = 1, 2, \ldots, N$ and $j = 0, 1, \ldots, k$.

Unconditional posterior for μ_t

Recall from step 4 that

$$(\mu_t \,|\, M_{t-k}^t, D_t) \sim \Gamma(y_t + r_t(i_{t-k}^t), \Delta_{i_t} + s_t(i_{t-k}^t)).$$

Proceeding as above

$$\mathbb{E}(\mu_t \,|\, D_t) = \sum_{i_t=1}^{N} \cdots \sum_{i_{t-k}=1}^{N} p_t(i_{t-k}^t) \frac{y_t + r_t(i_{t-k}^t)}{\Delta_{i_t} + s_t(i_{t-k}^t)} = h_t$$

and

$$\mathbb{V}\mathrm{ar}\,(\mu_t \,|\, D_t) = \sum_{i_t=1}^{N} \cdots \sum_{i_{t-k}=1}^{N} p_t(i_{t-k}^t) \Big(\frac{y_t + r_t(i_{t-k}^t)}{(\Delta_{i_t} + s_t(i_{t-k}^t))^2} + \Big(\frac{y_t + r_t(i_{t-k}^t)}{\Delta_{i_t} + s_t(i_{t-k}^t)} - h_t \Big)^2 \Big).$$

Unconditional posterior for $\boldsymbol{\theta}_t$

Step 8 gave that

$$(\boldsymbol{\theta}_t \,|\, M_{t-k+1}^t, D_t) \sim [\boldsymbol{m}_t(i_{t-k+1}^t), C_t(i_{t-k+1}^t)]$$

and calculations analogous to those in step 8 yield that

$$(\boldsymbol{\theta}_t \,|\, D_t) \sim [\boldsymbol{m}_t, C_t],$$

where

$$\boldsymbol{m}_t = \sum_{i_{t-k+1}=1}^{N} \cdots \sum_{i_t=1}^{N} p_t(i_{t-k+1}^t) \boldsymbol{m}_t(i_{t-k+1}^t)$$

and

$$C_t = \sum_{i_{t-k+1}=1}^{N} \cdots \sum_{i_t=1}^{N} p_t(i_{t-k+1}^t) \cdot (C_t(i_{t-k+1}^t) + (\boldsymbol{m}_t(i_{t-k+1}^t) - \boldsymbol{m}_t)^2).$$

3.3. Backward Filtering

In monitoring e.g. medical time-series the posterior model probabilities found in Section 3.1 will often be an intrinsic part of the surveillance procedure. These probabilities should be compared with the corresponding filtered marginal distributions of $(\mu_{t-j} \,|\, D_t)$. In general we therefore seek the first and second order moments of $(\boldsymbol{\theta}_{t-l} \,|\, D_t)$, where we will assume that $1 \leq l \leq k$. If we furthermore assume that $t \geq l + k$ no boundary value problems occur. The filtered marginal moments for $(\mu_{t-l} \,|\, D_t)$ may then be approximated by Taylor expanding $\exp(F_{t-l}^T \boldsymbol{\theta}_{t-l})$, see Appendix A.

4. Computational Issues

The above updating scheme has been implemented in the statistical software package R, see R Development Core Team (2005). Here one has to make special attention as long as $t \leq k$ since in that case the dimension of some of the above computations are different from the one stated.

Furthermore numerical problems arose in step 3 and 8 both of which will be addressed in the current section.

Numerical problems in step 3

Suppressing the indexes the likelihood function may be rewritten as

$$p(y \,|\, \cdot) = \begin{cases} \left(\frac{s}{\Delta+s}\right)^r & \text{if } y = 0 \\ \\ \Delta^y \Big(\prod_{i=1}^{y} \frac{y+r-i}{(y-(i-1))(\Delta+s)} \Big) \left(\frac{s}{\Delta+s}\right)^r & \text{if } y > 0. \end{cases}$$

However s and r may be quite large which courses R to round off $s/(\Delta + s)$ to 1 such that $(s/(\Delta + s))^r$ is set equal to 1 although it may be considerably smaller than 1. To overcome this note that

$$\left(\frac{s}{\Delta + s}\right)^r = \left(1 + \frac{\Delta}{s}\right)^{-r} = \left(1 + \frac{\Delta \exp(f)}{r}\right)^{-r},$$

where we have used the relations between r, s, f and q given in step 2. The limit

$$\lim_{r\to\infty} \left(1 + \frac{\Delta \exp(f)}{r}\right)^{-r} = \exp(-\Delta \exp(f))$$

proves useful whenever r is large since the right hand side causes no numerically problems in R.

We have computed the difference between $(1 + x/r)^{-r}$ and $\exp(-x)$ as a function of r for $x \in \{0.1, 0.5, 1, 3, 5, 7\}$ in R. Visual inspection revealed that the discrepancies for r

approximatively less than 10^4 was due to the poor approximation that $\exp(-x)$ provides for $(1 + x/r)^{-r}$. However, as r approaches 10^{12} the disagreement is caused by numerical instability in R when computing $(1 + x/r)^{-r}$. Therefore we have chosen to use the approximation, whenever $r > 10^7$.

Numerical problems in step 8

In calculating the posterior mean and variance for $\boldsymbol{\theta}_t$ division by $p_t(i_{t-k+1}^t)$ is required. Albeit, in theory, $p_t(i_{t-k+1}^t)$ is strictly positive, R may round it off to 0 making the division impossible. This will particularly be the case if the lag k is high and $(i_t, \ldots, i_{t-k+1})$ all correspond to non-steady states. In that case we have used $\pi(i_{t-k})$ as weights in step 8 instead of the correct $p_t(i_{t-k} \mid i_{t-k+1}^t)$.

5. Application

Mycoplasma pneumoniae is a frequent cause of respiratory infections in both children and adults worldwide, see Baum (2005). The type of infection diagnosed most frequently is pneumonia with an estimated annual incidence of one per 1000 persons; the incidence of non-pneumonic respiratory infections may be 10-20 times as high. Mycoplasma pneumoniae accounts for 1-8 percent of cases of pneumonia admitted to hospital in adulthood, see Bartlett and Mundy (1995), but far more cases are dealt with by general practitioners. Diagnostic tests include detection of specific antibodies and DNA-based methods such as polymerase chain reaction (PCR). In Denmark, Statens Serum Institut has offered a centralized diagnostic service for many years. This has revealed a distinctive picture of seasonal variation with a peak during the winter half-year and epidemics every 3 to 5 years, see Lind et al. (1997). However, other epidemiological patterns may occur, see Baum (2005). Antibiotics which normally are a first choice for patients with mild or moderate pneumonia may not provide coverage for Mycoplasma pneumoniae, and therefore a reliable and timely alert of an increased incidence of Mycoplasma pneumoniae would allow physicians in primary care as well as in hospitals to adjust empirical antibiotic therapy. The following data set (July 1994 to July 2005) represents samples of respiratory secretions submitted from all of Denmark to Statens Serum Institut for examination by PCR, see Jensen et al. (1989).

Let Y_t denote the total number of PCR-positive samples obtained at day t, $t = 1, 2, \ldots,$ 3132, where $t = 1$ corresponds to the 1st of January 1997. We then propose the following model:

$$\begin{aligned} Y_t \mid \mu_t, M_t(i_t) &\sim \text{Pois}(\mu_t \Delta_{i_t}) \\ \log(\mu_t) &= \lambda_t + \gamma_t \end{aligned}$$

where the trend λ_t is described as a linear growth with random perturbations in level and slope:

$$\begin{aligned} \lambda_t &= \lambda_{t-1} + \beta_t + \delta\lambda_t \\ \beta_t &= \beta_{t-1} + \delta\beta_t. \end{aligned}$$

The seasonal variation γ_t is a simple sine curve with period 365.25 and random perturbations

in amplitude, $\sqrt{a_t^2 + b_t^2}$, and phase, $-\arccos(a_t/\sqrt{a_t^2 + b_t^2})$:

$$\begin{aligned}\gamma_t &= a_t \sin(\phi_t) + b_t \cos(\phi_t)\\ a_t &= a_{t-1} + \delta a_t\\ b_t &= b_{t-1} + \delta b_t.\end{aligned}$$

Here $\phi_t = 2\pi t/365.25$. Let

$$\boldsymbol{\theta}_t^T = (\lambda_t, \beta_t, a_t, b_t).$$

Then the evolution equation can be written in matrix form as $\boldsymbol{\theta}_t = G\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$, where

$$G = \begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \quad \text{and} \quad \boldsymbol{\omega}_t = \begin{bmatrix}\delta\lambda_t + \delta\beta_t\\ \delta\beta_t\\ \delta a_t\\ \delta b_t\end{bmatrix}$$

and $\log(\mu_t) = F_t^T\boldsymbol{\theta}_t$ with

$$F_t^T = \begin{bmatrix}1 & 0 & \sin(\phi_t) & \cos(\phi_t)\end{bmatrix}.$$

The period from the 1st of July 1994 to the 31th of December 1996 was used as input to the analogous generalized linear model in order to obtain a bet for the prior mean $\boldsymbol{m}_0$. The prior variance C_0 is defined as a diagonal matrix with the square root of the diagonals being equal to

$$\begin{aligned}\sigma_{\text{prior}}(\lambda_0) &= 3 \cdot 10^{-5} & \sigma_{\text{prior}}(\beta_0) &= 4 \cdot 10^{-4}\\ \sigma_{\text{prior}}(a_0) &= 3 \cdot 10^{-5} & \sigma_{\text{prior}}(b_0) &= 3 \cdot 10^{-5}.\end{aligned}$$

We defined the following three model states: 1) steady state, 2) level and slope change and 3) outlier state, where the second is meant to represent an epidemic situation, in which the evolution shows erratic behaviour in level as well as in slope. Let $\sigma_j(X)$ correspond to the standard deviation of the random variable X when model j applies. The modelling of the three states was achieved by defining

$$\begin{aligned}\sigma_1(\delta\lambda_t) = \sigma_1(\delta\beta_t) &= 1 \cdot 10^{-7}\\ \sigma_1(\delta a_t) = \sigma_1(\delta b_t) &= 1 \cdot 10^{-6}\end{aligned}$$

and

$$\sigma_2(\delta\lambda_t) = \sigma_2(\delta\beta_t) = 0.5.$$

These values were all found empirically. Then

$$W_t(j) = \begin{bmatrix}\sigma_j^2(\delta\lambda_t) + \sigma_j^2(\delta\beta_t) & \sigma_j^2(\delta\beta_t) & 0 & 0\\ \sigma_j^2(\delta\beta_t) & \sigma_j^2(\delta\beta_t) & 0 & 0\\ 0 & 0 & \sigma_j^2(\delta a_t) & 0\\ 0 & 0 & 0 & \sigma_j^2(\delta b_t)\end{bmatrix}.$$

Finally we let

$$(\Delta_1, \Delta_2, \Delta_3) = (1, 1, 5)$$

corresponding to an outlier having a five-fold intensity compared with the current level. Furthermore

$$(\pi(1), \pi(2), \pi(3)) = (99.85\%, 0.1\%, 0.05\%)$$

implying a prior belief of approximately 3 level and slope changes and $1 - 2$ outliers in the time period analyzed.

Figure 2 and 3 show (a) the observed Mycoplasma counts (black vertical bars) and $\mu_t \,|\, D_{t+\text{lag}}$ (green line), (b) $\mathrm{P}(M_t(2) \,|\, D_{t+\text{lag}})$ and (c) $\mathrm{P}(M_t(3) \,|\, D_{t+\text{lag}})$ for lag $= 0, 1$, respectively. Clearly there has been a large epidemic in the winter 1998/1999 and a less prominent one in the winter 2004/2005. It is evident from Figure 2 that the multi-process

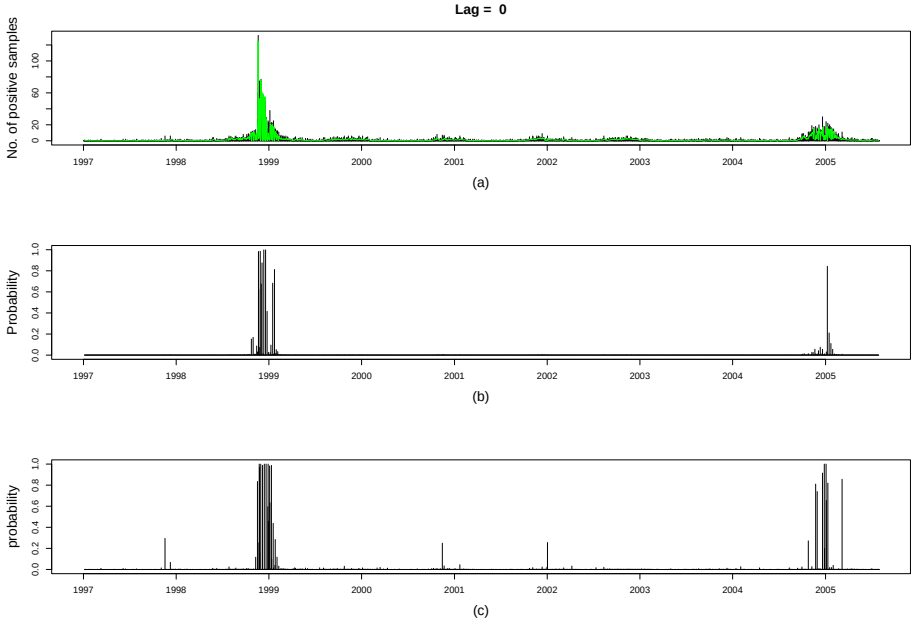


Figure 2. Mycoplasma data. (a) shows the observed Mycoplasma counts (black vertical bars) and $\mu_t \,|\, D_t$ (green line), (b) $\mathrm{P}(M_t(2) \,|\, D_t)$ and (c) $\mathrm{P}(M_t(3) \,|\, D_t)$.

model initially interprets the large counts observed in 1998/1999 and 2004/2005 as corresponding to either level and slope changes or outliers. On the other hand there seems to be no false positives in the very early prediction of an epidemic. Figure 3 reveals that looking just one day back in time greatly improves the discrimination between the epidemic and the outlier state and it is obvious that the model is able to recognize departures from the non-epidemic situation.

From Figure 3 it appears as if $(\mu_t \,|\, D_{t+1})$ alternates between high and low values during an epidemic. This might be accessed by exploring the one-step back distribution for the

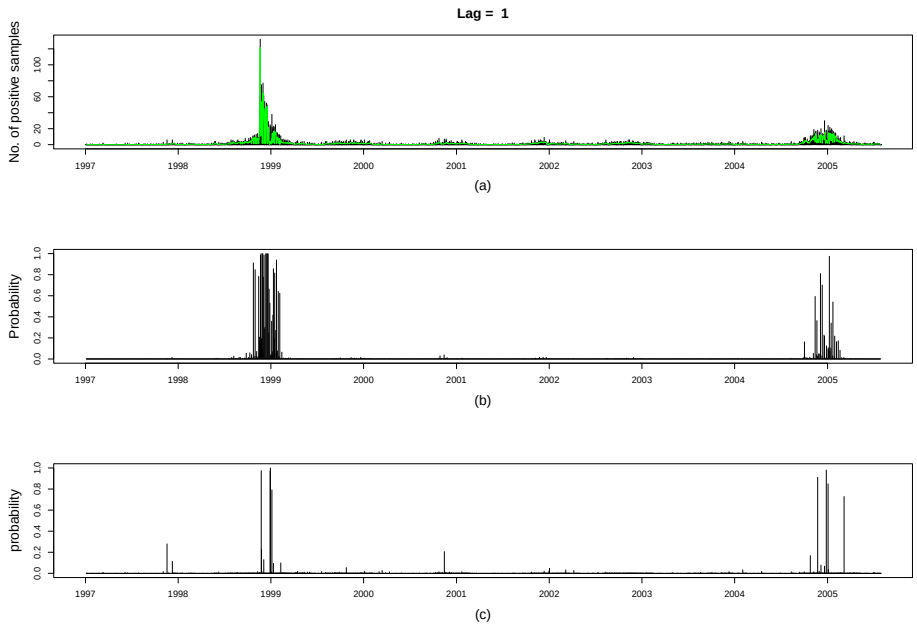


Figure 3. Mycoplasma data. (a) shows the observed Mycoplasma counts (black vertical bars) and $\mu_t \mid D_{t+1}$ (green line), (b) $\mathrm{P}(M_t(2) \mid D_{t+1})$ and (c) $\mathrm{P}(M_t(3) \mid D_{t+1})$.

intercept and slope. If ξ_t is the intercept that applies at time t then

$$\xi_t = \lambda_t - t \cdot \beta_t.$$

Figure 4(b) and 4(c) show a plot of $\mathbb{E}(\xi_t \mid D_{t+1})$ and $\mathbb{E}(\beta_t \mid D_{t+1})$ (Figure 4(a) is identical with Figure 3(a)). It is apparent from this figure that the slope – quite surprisingly – oscillates between positive and negative values during an epidemic. To assess the reason for this, Figure 5 provides the plots in Figure 4 restricted to the time period from the 1st of October 1998 to the 1st of Marts 1999. We see here that the negative slope is caused by a marked weekday variation due to the fact that hardly any samples are drawn in the weekends. Therefore an obvious improvement of the model would be to take this weekday variation into account. A more detailed analysis of the Mycoplasma data set – including the possibility for a day-to-day variation – will be published in a subsequent paper.

6. Discussion

Infectious diseases continue to pose a threat to society and it has been augmented by the continuous recognition of new pathogens, the potential for rapid dissemination around the globe, and the risk of bioterror. Only a limited set of infections is notifiable through a statutory surveillance system and a few including influenza may be the target of intensified surveillance at immediate contact with patients ('sentinel systems'). Both in hospitals and in the primary care setting surveillance needs resources for the collection of data. Often diagnoses must be confirmed by laboratory tests, and some delay is inherent in notification

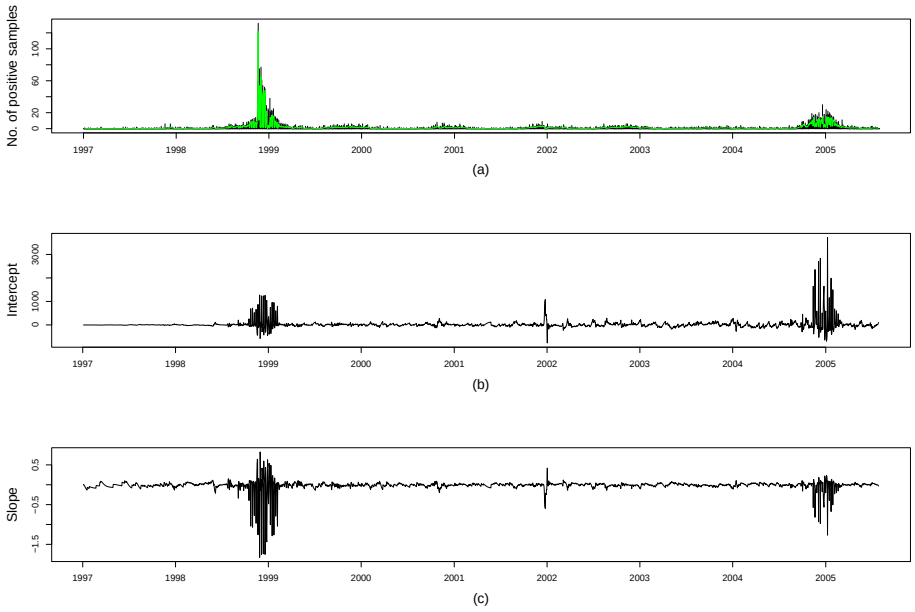


Figure 4. Mycoplasma data. (a) shows the observed Mycoplasma counts (black vertical bars) and $\mu_t \mid D_{t+1}$ (green line), (b) the evolution of the intercept $\mathbb{E}(\xi_t \mid D_{t+1})$ and (c) the slope $\mathbb{E}(\beta_t \mid D_{t+1})$.

procedures. Dependent on the scope of the surveillance system data are aggregated at regional, national or international level. Although numbers of observations may be large for some categories of infections others remain rare as exemplified by the nearly 2000 different types of Salmonella. Legionella infections is another example of a rare event even with data from the entire Europe, see Joseph (2004). For infections not being a target of formal surveillance systems useful information may still be available from health administrative systems such as hospital discharge registries and laboratory information systems. Health services automatically cumulate such data in dedicated databases with little delay and they pertain to the population served. Inherently, the numbers of observations will often be small.

Thus there are basically two applications of the proposed warning system: First of all it may support existing statutory surveillance systems especially in situations with rare occurrences. Secondly it may be embedded locally in the health administrative system where data are already available but where automated surveillance usually is not implemented. Thus the Poisson assumption of the presented model encompasses the handling of both of these surveillance situations. This implies a unified model approach in all situations, the Poisson assumption being crucial especially for surveillance of relatively rare events.

Furthermore the model is highly modular in the sense that separate models are formulated for the change-point component, the evolutional component and the observational component. It deserves notice that there is a conceptual difference between the change-point component and the evolutional component: In the change-point component the expected *frequency* of aberrant incidents is specified whereas the evolutional component describes the

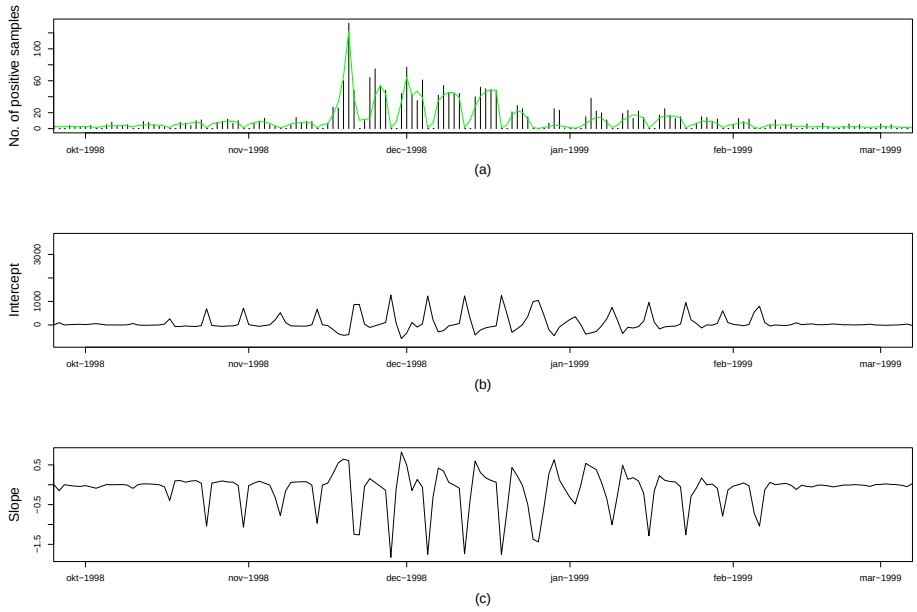


Figure 5. Mycoplasma data from the 1st of October 1998 to the 1st of Marts 1999. (a) shows the observed Mycoplasma counts (black vertical bars) and $\mu_t \mid D_{t+1}$ (green line), (b) the evolution of the intercept $\mathbb{E}(\xi_t \mid D_{t+1})$ and (c) the slope $\mathbb{E}(\beta_t \mid D_{t+1})$.

nature of the steady state behavior and the relevant aberrations. The evolutional component is itself modular as e.g. the seasonal variation, trend and other effects are modeled separately. This adds transparency and flexibility into the model specification. Moreover extensions and generalizations are easily adapted in this modular structure.

In the application we have focused on the posterior model probabilities, i.e. the probability of the occurrence of a recent change-point in the time series. Output on a probability scale has several advantages: Foremost it will be familiar to most of the expert users of the surveillance system. For specific applications the model may be integrated in a decision support system provided that dependable estimates of the associated costs of false positives and negatives can be obtained.

There is a versatility of other possible outputs from the proposed model. The posterior moments of the evolutional component contain information valuable for quality control, resource management purposes or epidemiological research. In laboratory information systems the former will primarily manifest itself in the surveillance of specific specimens, where unexpected changes in the base activity may stem from laboratory inconsistencies, see Dessau and Steenberg (1993). Examples of output relevant for resource management could be the contemporary growth rate or the expected number of counts within a specified time frame. Gradual changes in the pattern of seasonal variation, e.g. the peak-to-trough ratio or the time for peak may, on the other hand, be of general epidemiological interest, and can be valuable output for observational studies.

An issue yet to be solved is the calibration of such models. So far the applicability

of the model necessitate the disposal of learning data and the calibration is more or less obtained by trial-and-error. There is hence a need for the development of efficient parameter estimation methods.

Acknowledgements

We thank Anders Koch and Thomas Hjuler, Department of Epidemiology Research, Statens Serum Institut, Denmark for construction and advice on use of the mycoplasma database.

A. Backward Filtering

Input to the updating scheme at time t
Suppose that at time t the updating procedure described in Section 3.1 has been run and assume that

$$(\boldsymbol{\theta}_{t-l+1} \,|\, M_{t-k}^t, D_t) \sim [\boldsymbol{m}_t^{(l-1)}(i_{t-k}^t), C_t^{(l-1)}(i_{t-k}^t)] \tag{9}$$

is known and that $1 \leq l \leq k$.

l-step back filtering
The idea is to find the moments of $(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_t)$ and then use (9) to find $\boldsymbol{m}_t^{(l)}(i_{t-k}^t)$ and $C_t^{(l)}(i_{t-k}^t)$.

Now

$$p(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_t) \propto p(\boldsymbol{y} \,|\, \boldsymbol{\theta}_{t-l+1}, \boldsymbol{\theta}_{t-l}, M_{t-k}^t, D_{t-l}) p(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_{t-l}),$$

where $\boldsymbol{y} = (y_t, \ldots, y_{t-l+1})$. Knowing $(\boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_{t-l})$ it follows that $\boldsymbol{y}$ is independent of $\boldsymbol{\theta}_{t-l}$ and hence the above reduces to

$$\begin{aligned} p(\boldsymbol{y} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_{t-l}) p(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_{t-l}) & \\ \propto p(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_{t-l}) & \\ \propto p(\boldsymbol{\theta}_{t-l+1} \,|\, \boldsymbol{\theta}_{t-l}, M_{t-k}^t, D_{t-l}) p(\boldsymbol{\theta}_{t-l} \,|\, M_{t-k}^t, D_{t-l}). & \end{aligned}$$

Since $k \geq l$ the model at time $t-l+1$ appears in the first conditioning and hence the above is equal with

$$p(\boldsymbol{\theta}_{t-l+1} \,|\, \boldsymbol{\theta}_{t-l}, M_{t-l+1}) p(\boldsymbol{\theta}_{t-l} \,|\, M_{t-k}^{t-l}, D_{t-l}),$$

where we have used that $p(\boldsymbol{\theta}_{t-l} \,|\, M_{t-k}^t, D_{t-l})$ does not depend on the future.

The evolution equation in (6) implies that

$$(\boldsymbol{\theta}_{t-l+1} \,|\, \boldsymbol{\theta}_{t-l}, M_{t-l+1}) \sim [G_{t-l+1}\boldsymbol{\theta}_{t-l}, W_{t-l+1}(i_{t-l+1})].$$

Appropriately weighting the moments of $(\boldsymbol{\theta}_{t-l} \,|\, M_{t-l-k+1}^{t-l}, D_{t-l})$ (found in step 8) with $p_{t-l}(i_{t-l-k+1}^{t-k-1} \,|\, i_{t-k}^{t-l})$ (step 7) yields

$$(\boldsymbol{\theta}_{t-l} \,|\, M_{t-k}^{t-l} D_{t-l}) \sim [\boldsymbol{b}_{t-l}(i_{t-k}^{t-l}), B_{t-l}(i_{t-k}^{t-l})].$$

The fact that the above two distributions are only partially specified via their first and second order moments implies that the mean and variance of

$$(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_t)$$

are not well-defined. The mean and variance are therefore set equal to the linear Bayes estimates. Suppressing the subscripts these are given by

$$\mathbb{E}(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_t) = \boldsymbol{b}_{t-l} + B_{t-l} G_{t-l+1}^T U^{-1} (\boldsymbol{\theta}_{t-l+1} - G_{t-l+1} \boldsymbol{b}_{t-l}), \tag{10}$$

where

$$U = U(i_{t-k}^{t-l+1}) = G_{t-l+1} B_{t-l}(i_{t-k}^{t-l}) G_{t-l+1}^T + W_{t-l+1}(i_{t-l+1})$$

and

$$\mathbb{V}\mathrm{ar}\,(\boldsymbol{\theta}_{t-l} \,|\, \boldsymbol{\theta}_{t-l+1}, M_{t-k}^t, D_t) = B_{t-l} - B_{t-l} G_{t-l+1}^T U^{-1} G_{t-l+1} B_{t-l}. \tag{11}$$

Applying the above together with (9) and the relations $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \,|\, Y))$ and $\mathbb{V}\mathrm{ar}\,(X) = \mathbb{E}(\mathbb{V}\mathrm{ar}\,(X \,|\, Y)) + \mathbb{V}\mathrm{ar}\,(\mathbb{E}(X \,|\, Y))$ yields

$$(\boldsymbol{\theta}_{t-l} \,|\, M_{t-k}^t, D_t) \sim [\boldsymbol{m}_t^{(l)}(i_{t-k}^t), C_t^{(l)}(i_{t-k}^t)],$$

where

$$\boldsymbol{m}_t^{(l)}(i_{t-k}^t) = \boldsymbol{b}_{t-l}(i_{t-k}^{t-l}) + B_{t-l}(i_{t-k}^{t-l}) G_{t-l+1}^T U^{-1} (\boldsymbol{m}_t^{(l-1)}(i_{t-k}^t) - G_{t-l+1} \boldsymbol{b}_{t-l}(i_{t-k}^{t-l}))$$

and

$$\begin{aligned} C_t^{(l)}(i_{t-k}^t) = B_{t-l}(i_{t-k}^{t-l}) - B_{t-l}(i_{t-k}^{t-l}) G_{t-l+1}^T U^{-1} G_{t-l+1} B_{t-l}(i_{t-k}^{t-l}) + \\ B_{t-l}(i_{t-k}^{t-l}) G_{t-l+1}^T U^{-1} C_t^{(l-1)}(i_{t-k}^t) U^{-1} G_{t-l+1} B_{t-l}(i_{t-k}^{t-l}). \end{aligned}$$

This finally implies that

$$\mathbb{E}(\boldsymbol{\theta}_{t-l} \,|\, D_t) = \sum_{i_t=1}^{N} p_t(i_{t-k}^t) \boldsymbol{m}_t^{(l)}(i_{t-k}^t) = \boldsymbol{v}_t^{(l)}$$

and

$$\mathbb{V}\mathrm{ar}\,(\boldsymbol{\theta}_{t-l} \,|\, D_t) = \sum_{i_t=1}^{N} p_t(i_{t-k}^t) \cdot \left[C_t^{(l)}(i_{t-k}^t) + \left(\boldsymbol{m}_t^{(l)}(i_{t-k}^t) - \boldsymbol{v}_t^{(l)} \right)^2 \right] = V_t^{(l)}.$$

Therefore

$$\begin{aligned} \mathbb{E}(\eta_{t-l} \,|\, D_t) &= F_{t-l}^T \boldsymbol{v}_t^{(l)} = f_t^{(l)} \qquad \text{and} \\ \mathbb{V}\mathrm{ar}\,(\eta_{t-l} \,|\, D_t) &= F_{t-l}^T V_t^{(l)} F_{t-l} = q_t^{(l)}. \end{aligned}$$

Taylor expansion to the first order of $\mu_{t-l} = \exp(\eta_{t-l})$ around $f_t^{(l)}$ yields

$$\begin{aligned} \mathbb{E}(\mu_{t-l} \,|\, D_t) &\simeq \exp(f_t^{(l)}) \qquad \text{and} \\ \mathbb{V}\mathrm{ar}\,(\mu_{t-l} \,|\, D_t) &=\simeq \exp(2 f_t^{(l)}) q_t^{(l)}. \end{aligned}$$

Initialization
To initialize the above updating scheme we need the mean and variance of

$$(\boldsymbol{\theta}_t \mid M_{t-k}^t, D_t)$$

both of which were given in step 1 in Section 3.1.

References

Bartlett, J. and L. M. Mundy (1995, December). Community-acquired pneumonia. *New England Journal of Medicine 333*(24), 1618–1624.

Baum, S. G. (2005). Mycoplasma pneumoniae and atypical pneumonia. In G. Mandell, J. Bennett, and D. Raphael (Eds.), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, pp. 2271–2280. Elsevier Inc. Philadelphia PA.

Bolstad, W. M. (1995, March). The Multiprocess Dynamic Poisson Model. *Journal of the American Statistical Association 90*(429), 227–232.

Cooper, B. and M. Lipsitch (2004). The analysis of hospital infection data using hidden Markov models. *Biostatistics 5*(2), 223–237.

Dessau, R. B. and P. Steenberg (1993, April). Computerized Surveillance in Clinical Microbiology with Time Series Analysis. *Journal of Clinical Microbiology 31*(4), 857–860.

Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer-Verlag New York, Inc. ISBN 0-387-95054-0.

Farrington, C., N. Andrews, A. Beale, and M. Catchpole (1996). A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *Journal of the Royal Statistical Society. Series A 159*(3), 547–563.

Finkenstädt, B. F. and B. T. Grenfell (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Applied Statistics 49*(2), 187–205.

Gordon, K. and A. Smith (1990, June). Modelling and Monitoring Biomedical Times Series. *Journal of the American Statistical Association 85*(410), 328–337.

Harrison, P. and C. Stevens (1971, December). A Bayesian Approach to Short-term Forecasting. *Operational Research Quarterly 22*(4), 341–362.

Harrison, P. and C. Stevens (1976). Bayesian Forecasting. *Journal of the Royal Statistical Society. Series B 38*(3), 205–247.

Jensen, J. S., J. Søndergard-Andersen, S. A. Uldum, and K. Lind (1989, November). Detection of Mycoplasma pneumoniae in simulated clinical samples by polymerase chain reaction. Brief report. *Acta Pathologica, Microbiologica et Immunologica Scandinavica 97*(11), 1046–1048.

Joseph, C. A. (2004). Legionnaires' disease in Europe 2000-2002. *Epidemiology and Infection 132*(3), 417–424.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press Inc., New York. ISBN 0-19-852219-3.

Lind, K., M. W. Benzon, J. S. Jensen, and W. A. J. Clyde (1997, July). A seroepidemiological study of Mycoplasma pneumoniae infections in Denmark over the 50-year period 1946-1995. *European Journal of Epidemiology 13*(5), 581–586.

R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

Smith, A. and M. West (1983, December). Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter. *Biometrics 39*(4), 867–878.

SSI (2005). *Beregning af udbrudsstatus*. Statens Serum Institut, Denmark. http://www.ssi.dk/graphics/html/udbrudmonitor/siderne/Hovedindex.htm, (accessed 2005-12-19).

Strat, Y. L. and F. Carrat (1999). Monitoring epidemiologic surveillance data using hidden Markov models. *Statistics in Medicine 18*(24), 3463–3478.

West, M. (1986). Non-normal multi-process models. Technical Report Research Report 81, Department of Statistics, University of Warwick.

West, M. and J. Harrison (1999). *Bayesian Forecasting and Dynamic Models*. Springer. ISBN 0-387-94725-6.

Whittaker, J. and S. Frühwirth-Schnatter (1994). A Dynamic Changepoint Model for Detecting the Onset of Growth in Bacteriological Infections. *Applied Statistics 43*(4), 625–640.