STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:3051–3069 (DOI: 10.1002/sim.963)

Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology

Flávio Fonseca Nobre[1], Ana Beatriz Soares Monteiro[2], Paulo Roberto Telles[3], and G. David Williamson[4,*,†]

[1]*Programa de Engenharia Biomédica - COPPE/UFRJ, Cidade Universitária - Ilha do Fundão, P.O. Box 68510, 21945-970 - Rio de Janeiro - RJ, Brazil*
[2]*Departamento de Estatística - GET/EGM/CEG - UFF, Rua Mário Santos Braga s/no - Praça do Valonguinho, 24020-110 - Centro - Niterói - Rio de Janeiro, Brazil*
[3]*NEPAD/UERJ - State University of Rio de Janeiro, Rua Fonseca Teles 121, 4 andar, 20940-200 - Rio de Janeiro - RJ, Brazil*
[4]*Epidemiology Program Office, MS K74, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A.*

SUMMARY

One goal of a public health surveillance system is to provide a reliable forecast of epidemiological time series. This paper describes a study that used data collected through a national public health surveillance system in the United States to evaluate and compare the performances of a seasonal autoregressive integrated moving average (SARIMA) and a dynamic linear model (DLM) for estimating case occurrence of two notifiable diseases. The comparison uses reported cases of malaria and hepatitis A from January 1980 to June 1995 for the United States. The residuals for both predictor models show that they were adequate tools for use in epidemiological surveillance. Qualitative aspects were considered for both models to improve the comparison of their usefulness in public health. Our comparison found that the two forecasting modelling techniques (SARIMA and DLM) are comparable when long historical data are available (at least 52 reporting periods). However, the DLM approach has some advantages, such as being more easily applied to different types of time series and not requiring a new cycle of identification and modelling when new data become available. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

Public health surveillance represents an important means for continuously collecting, analysing, interpreting and disseminating health data essential to prevention and control efforts [1]. One use of a public health surveillance system is to facilitate the detection of unusual patterns in the occurrence of diseases and other health events. Toward this goal, different

*Correspondence to: G. David Williamson, Epidemiology Program Office, MS K74, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341, U.S.A.
†E-mail: dxw2@.cdc.gov

Contract/grant sponsor: MCT/FINEP/CNPq (Brazil)

Received November 1998
Accepted February 2001

analytical methods have been explored to forecast expected numbers of reported cases and compare with the observed numbers of cases. Some of these approaches assume that surveillance reports from successive time periods are independent and may also require knowledge of the probability distribution underlying the data [2]. However, surveillance data are expected to exhibit temporal dependence, suggesting that time-series techniques might be more appropriate for the analysis and forecasting of expected numbers of reported cases.

Analysis of time-series data has long been of interest in the literature, in particular the use of the Box–Jenkins approach also known as autoregressive integrated moving average (ARIMA) models [3]. To account for the temporal dependence of surveillance data, throughout the 1980s there was a marked increase in the use of these models in epidemiology. Time-series studies have focused on mortality due to pneumonia and influenza to forecast mortality or to calculate excess mortality, using a univariate approach, [4, 5], or using a multivariate focus [6] to improve mortality estimates. Time-series methods have also been used extensively to study infectious diseases [7], including the proposal of a method to define an alert threshold for supporting public health surveillance [8]. More recently, Williamson and Weatherby [9] report a method combining ARIMA models and statistical process control charts to model public health surveillance data and produce a signal when data aberrations are detected.

Other forecasting techniques applied to monitor changes in public health surveillance data involve the method of exponential smoothing [10]. Using the exponential smoothing model for forecasting and properties of numerical derivatives a method was developed for detecting points at which the reporting of diseases changes [11]. Another method explored for forecasting and detecting abnormal patterns in surveillance data was based on the Kalman filter [12]. This approach provides a flexible framework to analyse time series in an adaptive form, and has been recently applied for monitoring AIDS surveillance data [13].

A class of models that seems to be very useful as general purpose models for forecasting is the dynamic linear model (DLM) [14]. This modelling approach assumes the parameters as having prior distributions, independent of the data, and the data are used to update this distribution. Both the Kalman filter method and the DLM approach are in essence Bayesian methods, since their updating scheme is based on the Bayes theorem, and both allow the inclusion of subjective information, such as an expert opinion, for the construction of the *a priori* distribution.

The problem of choosing a forecast approach depends on the relative performance of the models for monitoring and prediction, with an adequate interpretation of the phenomenon under study. Comparison of different forecasting models with available surveillance data can help health workers to understand limitations of their data and may facilitate the selection of a time series model for forecasting. The objectives of the present paper are to compare DLM and SARIMA models for use on epidemiological time series and evaluate the forecasting performance and other qualitative aspects. This comparison may be helpful for the epidemiologist to choose the most suitable methodology in a given situation.

We outline the SARIMA and DLM time series models in Section 2, including description of the quantitative measures used for comparing the models and the data used for the comparison process. The results of modelling the time series are shown in Section 3. In Section 4 we compare the results of the proposed models. A general discussion and some suggestions for future research are presented in Sections 5 and 6.

2. METHODS

The Box–Jenkins methodology and the dynamic linear model (DLM) will be briefly described here. For a more detailed presentation of the Box–Jenkins approach in health surveillance, see Helfenstein [7]. Since the DLM approach is less known, we provide a more extensive description of this method.

2.1. Seasonal Autoregressive Integrated Moving Average Models (SARIMA)

The seasonal model applied in this research is the most general form of a univariate class of models originally presented by Box and Jenkins [3]. It has been extensively studied and used in different fields, such as economy, industry and, more recently, in public health. This model building process is designed to take advantage of the association in the sequentially lagged relationships that usually exist in data collected periodically.

An important concept for the model building process is that of stationarity, which implies that the probabilistic structure of the series does not change with time. However, very often epidemiological time series in a public health surveillance system have a trend component, that is, they have no fixed mean. It has been found that removal of trends in the mean can usually be achieved by differencing the time series. In this investigation, we modelled the trend in the data with difference equations [3]. A first-order difference of the series z_t is given by w_t, the difference between points in the series one unit apart, calculated as $w_t = z_t - z_{t-1}$. We may also write w_t in terms of the backward shift operator B as $w_t = (1-B)z_t$, and so the dth order differencing is obtained as $(1-B)^d z_t$.

Besides trend components, epidemiological time series may exhibit seasonality. Box and Jenkins have extended the above idea of ordinary differencing by forming seasonal differences $w_t = z_t - z_{t-s} = (1-B^s)z_t$, where s is the seasonal period of the data, say 12 months. Therefore, the most general Box–Jenkins model, the seasonal autoregressive integrated moving average (SARIMA), has the following form:

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d z_t = \theta(B)\Theta(B^s)a_t \tag{1}$$

with

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

$$\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$$

$$\Theta(B^S) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ}$$

where p is the autoregressive order, q the moving average order, d is the number of differencing operations, and P, D and Q are the corresponding seasonal orders.

After removal of trend and seasonal components, the process of model fitting involves identification, parameter estimation and diagnostic validation. At the identification stage, a tentative autoregressive moving average (ARMA) process is developed based on the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF). The shape of the ACF and PACF of the real epidemiological time series is compared with the

shape of the theoretical model. This process of comparison allows the definition of p and q, the orders of the ARMA model.

Having specified an initial model, the parameters of the candidate model are estimated. The last stage, diagnostic checking, examines the residuals to determine the adequacy of the model, or how well the residuals are randomly distributed about zero. If the model is judged adequate, it is used in the forecasting stage. If the fitted model is not adequate, the process of identification, estimate and diagnostic checking is repeated until a satisfactory model is found.

2.2. Dynamic linear Bayesian models (DLM)

The dynamic Bayesian model, formalized by West and Harrison [14], is based on the premise that uncertainties must be represented by a probability distribution. The basic approach of DLM modelling comprises defining a structural model for the time series, a probabilistic representation of information about all parameters and observable information and inference and forecasting derived by summarizing appropriate posterior and predictive probability distributions.

When a trend and a seasonal component compose an epidemiological time series, we let z_t denote the time series at time t, and we can write the structural model as

$$z_t = \mu_t + \psi_t + \varepsilon_t \tag{2}$$

where μ_t is the stochastic trend, ψ_t is a seasonal component around the long-run trend, and ε_t is a white noise component representing the observational errors. The trend component μ_t is obtained using

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{t,1} \\ \beta_t &= \beta_{t-1} + \omega_{t,2} \end{aligned} \tag{3}$$

where β_t is the slope of the trend and $\omega_{t,1}$ and $\omega_{t,2}$ are random errors. A seasonal component of the structural model can be described using the Fourier series decomposition in the following form:

$$\psi_t = \sum_{i=1}^{k} [A_{i,t} \cos(2\pi f_i t) + B_{i,t} \sin(2\pi f_i t)] \tag{4}$$

where f_i, $i = 1, 2, \ldots k$, are the frequency in cycles per unit of time. Recalling that frequency is the inverse of time, we have that $f_1 = 1/s$; s being the seasonal component of the data, and for $i = 2, 3, \ldots, k$ we have the harmonics of the fundamental frequency f_1. The coefficients $A_{i,t}$ and $B_{i,t}$ are assumed to be time-variables, so that the model can handle non-stationary seasonality.

Once the basic structures for the components of the model are defined, assembling these into an aggregate form is possible, expressing the model as a dynamic linear model (DLM)

$$z_t = \boldsymbol{F}'\theta_t + v_t \qquad v_t \sim \mathrm{N}[0, \boldsymbol{V}_t] \tag{5}$$

$$\theta_t = \boldsymbol{G}\theta_{t-1} + \omega_t \qquad \omega_t \sim \mathrm{N}[0, \boldsymbol{W}_t] \tag{6}$$

where θ_t is the $(2k+2)\times 1$ vector of system parameters at time t and defined as $\theta_t' = (\mu_t, \beta_t, A_{i,t}, B_{i,t})$. $\boldsymbol{F}'$ is a vector of known constants and the observation errors, v_t, are assumed normally distributed with variance $\boldsymbol{V}_t$.

Equation (6) is the evolution or system equation describing the process of variation of the system parameters along with time, including their time-varying dependence. The time variation of the parameters is identified by the evolution or transition matrix, $\boldsymbol{G}$, and the evolution errors, ω_t, representing stochastic changes in the state vector, are assumed normally distributed with zero mean and known variance matrix $\boldsymbol{W}_t$.

Once specified, the model is implemented by updating priors to obtain posteriors using a sequential approach. At time t the prior information on the state $(\theta_t \mid D_{t-1})$ is obtained using the evolution equation where D_{t-1} denotes all the information that is available before observing the current value of the time series, z_t. The prior information is used in the observation equation to generate the model forecasts. The *posterior* distribution is then obtained through the Bayes theorem

$$P(\theta_t \mid D_t) \propto P(\theta_t \mid D_{t-1}) P(z_t \mid \theta_t, D_{t-1}) \tag{7}$$

where $P(z_t \mid \theta_t, D_{t-1})$ is the likelihood and $D_t = \{D_{t-1}, z_t\}$. Using normal prior and likelihood distributions, DLM produces $(\theta_t \mid D_t) \sim \mathrm{N}(m_t, C_t)$, where m_t and C_t are obtained by taking expectations on the evolution equation (6), that is, $m_t = \boldsymbol{G} m_{t-1}$ and $C_t = \boldsymbol{G} C_{t-1} \boldsymbol{G}^{\mathrm{T}} + \boldsymbol{W}_t$.

To start the DLM modelling process, it is necessary to specify the prior $(\theta_1 \mid D_0)$ for the state vector, before arrival of the first value of the time series. The usual procedure is to use an uninformative distribution, that is, a normal distribution with mean equal to zero and a large variance.

The observational variance V_t can be estimated using a mathematical relation with the expected value of the observed time series [14]. A special case is the power law function defined as $V_t = \phi^{-1} \mu_t^b$. For $b = 1$ the variance is proportional to the expected value (as a Poisson distribution). In the normal DLM, the precision parameter ϕ follows a gamma distribution with parameters estimated at each time. The elements of the variance–covariance matrix $\boldsymbol{W}_t$ can also be addressed using an *ad hoc* way, as $\boldsymbol{W}_t = [(1-\delta)/\delta]\boldsymbol{C}_{t-1,}$ where $\boldsymbol{C}_{t-1}$ is the posterior covariance matrix of the parameters at the previous time $t-1$ and δ is a discount factor $(0 < \delta \leqslant 1)$. The discount factor determines the contribution of past information on the current estimate of the parameters and it can be set at different values for each of the model components (trend and seasonality). For δ close to zero, current estimates are strongly dependent on recent information and for $\delta = 1$ the estimates are similar to the average of all the data, portraying a static model.

Until recently, the approach described above for DLM modelling was the only one available and is the one implemented in the BATS [15] software used here. Advances in computation have begun to extend the possibilities for exploiting other options for modelling the elements $\boldsymbol{V}_t, \boldsymbol{W}_t, \boldsymbol{F}'$ and $\boldsymbol{G}$ through the use of new MCMC methods [16].

The adequacy of the DLM model is evaluated using the logarithm of the predictive likelihood (LPL) [15]

$$\mathrm{LPL} = \sum_{t=1}^{T} \log[P(z_t \mid z_1, z_2, \ldots, z_{t-1}, D_0)]$$

where $P(\,)$ is the predictive distribution and D_0 is the prior information for time $t=1$. The larger the value of LPL, the better the fit of the model to the time series.

The fitting of the models is evaluated using quantitative measures and graphical features of the model. Three error measures are considered here [15]:

(i) mean square error

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T} e_t^2$$

(ii) mean absolute error

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T} |e_t|$$

(iii) mean absolute percentage error

$$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \frac{|e_t|}{z_t}$$

where e_t is the forecast error and T is the number of observations minus the number of parameters of the model. The larger the values of these measures, the worse will be the fit of the model to the time series.

Before describing the performance of the models on the selected time series, we present some qualitative comparison of the two forecasting methods. First, the Box–Jenkins methodology is well known, and there are several recent applications of its use in epidemiology, including its combination with statistical process charts for detecting aberrations [9]. DLM, on the other hand, is a methodology formalized in the late 1980s and has not yet been sufficiently explored for use in a broad array of applied sciences, including epidemiology. Second, several statistical packages incorporate the SARIMA modelling approach, yielding outputs that are sufficient for identification and estimating purposes. For DLM there is only a handful of specific software (for example, BATS) [15]. Though it has a simple structure, the DLM forecasting approach requires the specification of several parameters that are not easily understood and involves more complex analysis of the outputs from the user for identification. However, one advantage of the DLM method is that it does not require historical data and can be set up based on available expert opinion before data collection starts. This can be achieved by obtaining from the expert relevant information about estimated level of case occurrence and variability that then helps the forecaster structure the beliefs in a parametric model defining the initial prior probability density function.

Third, while SARIMA models are composed of polynomials with constant coefficients, DLM considers the parameters as random variables with known probability distributions. For the DLM framework, the parameters can be assumed as time-variables, introducing the possibility of incorporating in the model time-dependent factors, such as delays and variations in reporting case occurrences. Fourth, for the DLM approach the assumption of stationarity is not a prerequisite, as for the SARIMA modelling that often requires the use of differencing and transformations to stabilize the time series moments. Sometimes, disease series cannot be modified to achieve stationarity, thus negating the option of using Box–Jenkins methods. Fifth, for SARIMA modelling assuming a normal distribution for the model errors to simplify

inferences about the parameters is common. The use of Gaussian models is not a restriction for the DLM model because it allows any other distribution to be used. Sixth, generally SARIMA methods cannot produce viable models in the presence of sparse or missing data in the time series. In the presence of sparse or missing data, DLM increases the uncertainty of a parameter's distribution until data become available.

Finally, public health surveillance is an ongoing process, and adapting to changes as new data become available is not trivial when using the SARIMA framework. This method requires a new modelling effort as changes occur in the epidemiology, case definition or reporting of a disease or health event. For DLM modelling, the inclusion of a monitoring scheme to identify changes of patterns and adapting to the changes are easily set up through discount factors. This approach also allows the use of subjective forward interventions, such as an expert opinion, about future patterns of the time series.

2.3. Data sources

Reported number of cases of hepatitis A and malaria for the United States were used for comparing SARIMA and DLM forecasting approaches. Here data were collected by the Centers for Disease Control and Prevention (CDC) through their National Notifiable Diseases Surveillance System (NNDSS). Following previous work [2, 9], reduction of variability induced by other factors unrelated to the disease process was obtained by aggregating disease reports over a four-week period, which will be called a 'month'. Using this approach, each year results in 13 months. Figure 1 depicts a plot of these reported cases for the period January 1980 (1980.01) to June 1995 (1995.06). This figure shows that both disease time series are non-stationary. Hepatitis A displays a long-term trend, but not an expressive seasonality, whereas malaria shows a marked seasonal pattern, except the first year where a decreasing trend is present in the data.

Our main interest in this paper is to analyse the predictive performance of the models for several forecasted periods. The forecast period of 1989.01–1995.06 was chosen and divided into six periods of 13 months (corresponding to the six years of complete data reporting, 1989–1994) and one period with 6 months (corresponding to the period of data reports available for 1995).

3. TIME SERIES MODELLING RESULTS

3.1. SARIMA modelling

Initially, both time series (for hepatitis A and malaria) were analysed for identification of possible non-homogeneous variances. A practical tool for this analysis is the mean interquartile range plot: the interquartile range is plotted against the mean for each seasonal period. If the range is independent of the mean, no transformation is needed. If the plot displays a random scatter about a straight line, the logarithmic transformation is suggested. In our case it is seen in Figures 2(a) and (b) that the range was approximately independent of the mean, suggesting that there was no need of a variance-stabilizing transformation.

After identifying that there was no need of transformation, the steps of model identification, parameter estimation and diagnostic checking were done. For each modelling time interval

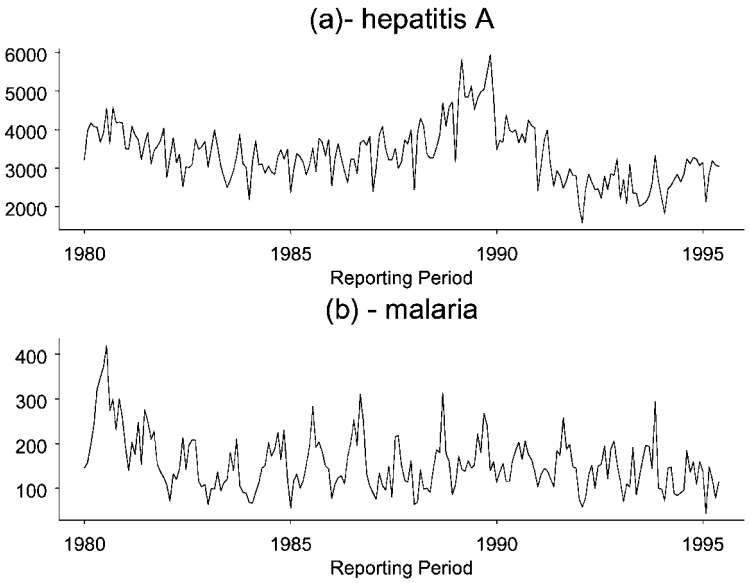

Figure 1. Monthly notifications for (*a*) hepatits A and (*b*) malaria for the U.S.A. from January 1980 to June 1995.

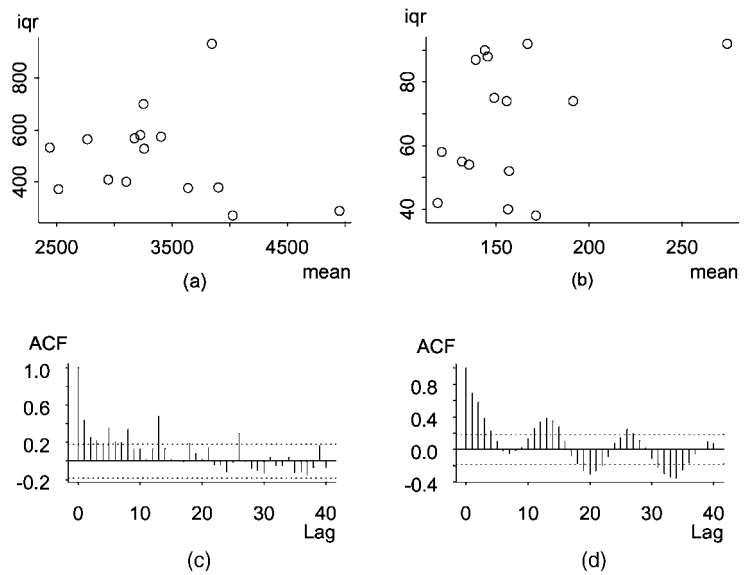

Figure 2. Mean interquartile range plot (excluding the six months of 1995) and estimated autocorrelation function for the period 1980.01–1988.13. (*a*) hepatitis A; (*b*) malaria; (*c*) hepatitis A; (*d*) malaria.

(1980.01 to 19*nn*.13, $nn = 88, 89, \ldots, 94$), we evaluated the pattern of the time series and the ACF to identify the degree of differencing to remove trend and seasonality.

The estimated ACF for both time series in the first interval (1980.01–1988.13) are depicted in Figures 2(c) and (d). For hepatitis A, the presence of a non-stationary mean is clearly visible in the time series plot and in the estimated ACF. A similar pattern for all intervals was obtained, exhibiting a periodicity of length 13, suggesting therefore a seasonal difference operation followed by a first-order differencing to induce stationarity. A strong seasonal component is present in all modelling periods of the malaria series. The estimated ACF for all intervals displays a significant autocorrelation coefficient at lag 13 and a slow decaying pattern, suggesting the need of a seasonal and first-order differencing operation. Additional support for the seasonal component is provided by statistical significant seasonal coefficients for all estimated models.

Once stationarity was achieved, both the ACF and PACF (Figure 3) were used to estimate the order of the models and the parameters were estimated using maximum likelihood method. We assessed model fit by a process called diagnostic checking that uses the residuals from the fitted model (residual = data − model). Figure 3 also depict the ACF of the residuals, showing that no structure is left. For the other time periods, the plot of the residual ACF also resembled those of a white noise.

For the ARIMA modelling, hepatitis A and malaria series produced the same model structure. Both required one seasonal and one non-seasonal difference and moving average terms. This structure is constant for all modelling periods. The estimated models for each period are presented in Tables I(a) and (b). They provide a parsimonious and reasonable representation of the data.

The estimated models depicted in Table I support the adequacy and stability of the models. It can be observed that little change occurred in the estimated model coefficients from the first period, 1980.01–1988.13, to the last period, 1980.01–1994.13. All estimated coefficients are statistically significant and satisfy Box–Jenkins invertibility conditions. Examination of the estimated residuals and their ACF and PACF did not show the presence of stochastic structure. The normal probability plot and histogram show an approximately normal distribution. These results, with the forecasting error measures (as given in Section 4), suggest that the estimated SARIMA models are adequate for forecasting.

3.2. DLM modelling

Given the evidence of seasonal and trend patterns of both hepatitis A and malaria, an adequate structural time series model for all data sets is the seasonal dynamic linear growth model. Although the trend component of malaria and the seasonal pattern of hepatitis A are not very strong, these structural components are modelled and their significance is evaluated during the estimating process.

Once the structural components of a time series are defined, the usual approach is to specify a set of values for the power law coefficient (b) and for the discount factor of each model component and evaluate the competitive model's performance [14]. For fitting performance evaluation, we specify the following values:

1. Discount factor for trend: 0.80, 0.85, 0.90 and 0.95.
2. Discount factor for seasonality: 0.90, 0.95, 0.98 and 1.00.

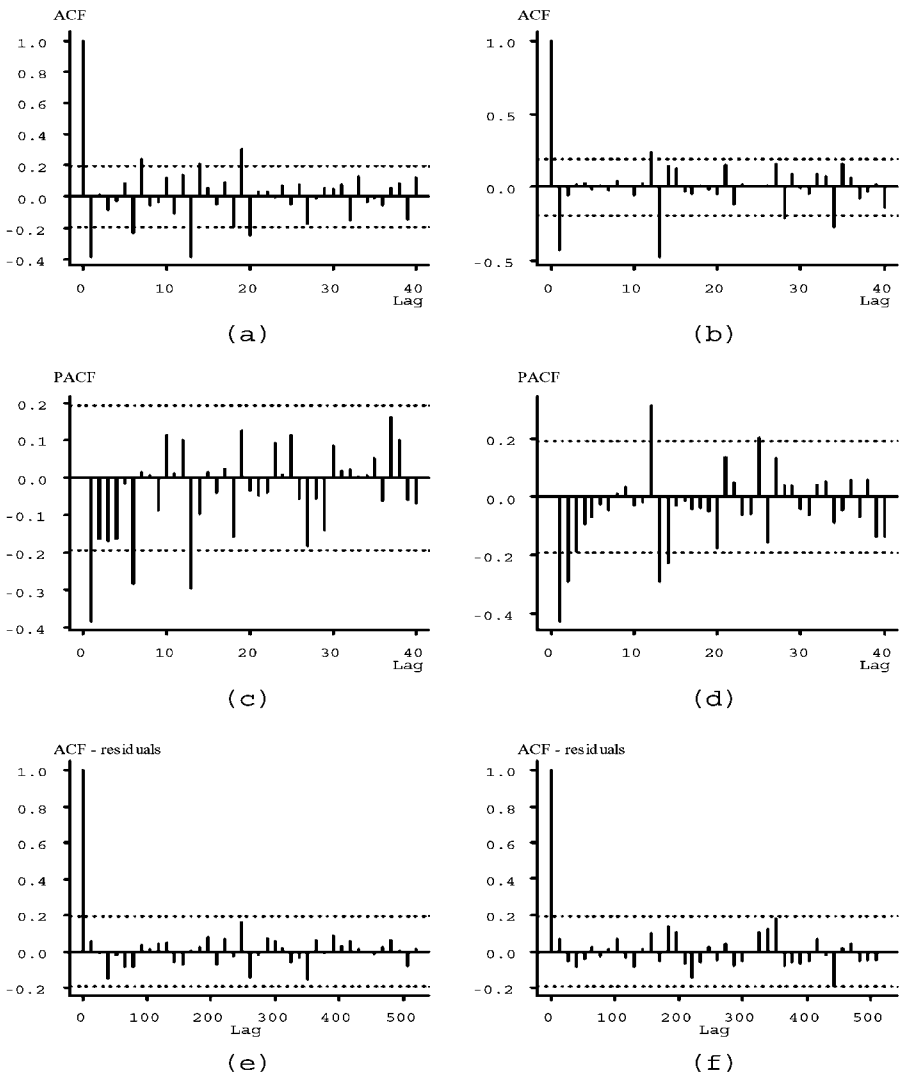

Figure 3. Estimated autocorrelation (ACF) and partial autocorrelaton (PACF) after removal of trends in the means and seasonality of both time series. Also, the ACF for the residuals is shown. (*a*), (*c*) and (*e*) hepatitis A; (*b*), (*d*) and (*f*) malaria.

3. Discount factor for precision (ϕ): 0.95, 0.99 and 1.00.
4. Power law coefficient (b): 0, 1 and 2.

The use of lower discount factors for trend and seasonality suggests that we have a lower belief in the priors, resulting in larger revised posterior variances. The use of larger discount factors says that we have a strong belief in our initial model (prior distribution). Use of a discount factor for the precision parameter reflects changes in the variance, possibly related

Table I. Estimated SARIMA models for hepatitis A and malaria.

Fitting period	Model
(*a*) *Hepatities A*	
1980.01–1988.13	$(1-B)(1-B^{13})Y_t = (1-0.58B)(1-0.66B^{13})a_t$
1980.01–1989.13	$(1-B)(1-B^{13})Y_t = (1-0.61B)(1-0.62B^{13})a_t$
1980.01–1990.13	$(1-B)(1-B^{13})Y_t = (1-0.49B)(1-0.72B^{13})a_t$
1980.01–1991.13	$(1-B)(1-B^{13})Y_t = (1-0.47B)(1-0.72B^{13})a_t$
1980.01–1992.13	$(1-B)(1-B^{13})Y_t = (1-0.50B)(1-0.71B^{13})a_t$
1980.01–1993.13	$(1-B)(1-B^{13})Y_t = (1-0.52B)(1-0.72B^{13})a_t$
1980.01–1994.13	$(1-B)(1-B^{13})Y_t = (1-0.52B)(1-0.71B^{13})a_t$
(*b*) *Malaria*	
1980.01–1988.13	$(1-B)(1-B^{13})Y_t = (1-0.63B)(1-0.60B^{13})a_t$
1980.01–1989.13	$(1-B)(1-B^{13})Y_t = (1-0.64B)(1-0.60B^{13})a_t$
1980.01–1990.13	$(1-B)(1-B^{13})Y_t = (1-0.64B)(1-0.61B^{13})a_t$
1980.01–1991.13	$(1-B)(1-B^{13})Y_t = (1-0.66B)(1-0.61B^{13})a_t$
1980.01–1992.13	$(1-B)(1-B^{13})Y_t = (1-0.68B)(1-0.60B^{13})a_t$
1980.01–1993.13	$(1-B)(1-B^{13})Y_t = (1-0.71B)(1-0.57B^{13})a_t$
1980.01–1994.13	$(1-B)(1-B^{13})Y_t = (1-0.71B)(1-0.61B^{13})a_t$

to truncations, changes in data recording and other unexplained errors. For the power law coefficients, since power b is directly related to the power used in the Box–Cox transformation, we use the above values corresponding to if our observational variance is constant or if there is a need of applying a transformation (that is, $b=0$ suggests no transformation, $b=1$ a square root transformation and $b=2$ a logarithm transformation).

DLM provides a monitoring procedure that can be used to compare, at each time, the proposed model with an alternative with the same predictive distribution, but with a wider variance. The alternative model is used for prediction at times where either an outlier or a structural change in the model is detected. At these times, the parameters' variance is increased so that most recent observations influence the updating of the posterior distribution more [15]. The monitoring involves the detection of upward or downward changes in the level and detection of non-specific changes related with the precision of the parameter distribution. Using discounted models, this can be done through an intervention by decreasing the discount factors. In the present application changing the discount factor for trend and seasonality to 0.10 was common and for precision to 0.90, facilitating the model's adaptation. The usefulness of this monitoring is verified comparing the performance of monitored versus non-monitored models.

Selection of the models was based on the logarithm of the predictive likelihood error (LPL) [14, 15]. The results of the modelling process are shown in Table II. The last column displays for each period the month that the monitors detect some form of level instability and change the corresponding discount factor. For the time series being modelled, there was no detection of changes in the precision factor (ϕ).

For both series, the fitted models are similar for all the modelling intervals. The power parameter, b, for all cases is equal to 1, suggesting that a more adequate approach is to use an observational variance proportional to the expected values of the series. This contrasts with the indication previously obtained using the mean-range plot for the SARIMA modelling.

Table II. Fitting results for hepatitis A and malaria.

Modelling interval	Discounts			Coefficient b	Intervention periods
	Trend	Season	Precision		
(*a*) *Hepatitis A*					
1980.01–1988.13	0.90	0.95	0.99	1	1988.10
1980.01–1989.13	0.90	0.95	0.99	1	1988.10
1980.01–1990.13	0.90	0.95	0.99	1	1988.10, 1990.02
1980.01–1991.13	0.80	0.95	0.99	1	1990.02
1980.01–1992.13	0.80	1.00	0.99	1	1982.09, 1987.04, 1989.03, 1992.02
1980.01–1993.13	0.80	0.99	0.99	1	1990.02, 1993.03
1980.01–1994.13	0.80	1.00	0.99	1	1982.09, 1987.04, 1989.03, 1992.02
(*b*) *Malaria*					
1980.01–1988.13	0.90	1.00	0.99	1	1984.12, 1988.10
1980.01–1989.13	0.90	1.00	1.00	1	1984.12, 1988.10
1980.01–1990.13	0.90	1.00	1.00	1	1984.12, 1988.10
1980.01–1991.13	0.85	1.00	0.99	1	1988.10
1980.01–1992.13	0.90	1.00	1.00	1	1984.12, 1988.10
1980.01–1993.13	0.90	1.00	1.00	1	1984.12, 1988.10, 1993.02
1980.01–1994.13	0.95	1.00	1.00	1	1984.12, 1988.10, 1993.12

Computing the increase in LPL error of the models with b equal 0, we observe that they are only marginally higher, being in the range of 0.08 per cent to 2.26 per cent for hepatitis A, and 0.5 per cent to 3.11 per cent for malaria.

Careful examination of the residuals did not show the presence of structure and, combined with results in Section 4 for the forecasting measures, suggest that the models are adequate.

4. COMPARISON OF FORECASTING PERFORMANCE

To evaluate the forecasting performance of SARIMA and DLM approaches, three forecasting error measures are used, MSE, MAE and MAPE. For each method and each modelling time period, a 13-step-ahead forecasting is obtained, except the last year that has only six available observations.

Results of the three forecasting error criteria for the two modelling approaches are shown in Table III. Some tentative conclusions can be drawn from these tables, which show that no modelling approach dominates the other:

(i) The DLM has smaller forecasting errors for four forecasted years of hepatitis A (1989, 1991, 1994 and 1995) and three forecasted years for malaria (1991, 1994 and 1995) suggesting a better performance in these cases.

(ii) The SARIMA model produces smaller errors for three forecasted years of hepatitis A (1990, 1992 and 1993) and three for malaria (1989, 1990 and 1993) showing a better performance in these cases.

(iii) For one forecasted year of malaria (1992), the results are not conclusive. In terms of the MSE, DLM performs better; for the MAPE criteria the SARIMA model produces smaller forecasted errors; and the evidence provided by MAE is that both have the same performance.

Table III. Forecasting performance for hepatitis A and malaria.*

Forecasting period	MSE ($\times 10^6$)		MAE		MAPE	
	SARIMA	DLM	SARIMA	DLM	SARIMA	DLM
(*a*) *Hepatitis A*						
1989	1.556	**0.698**	1106	**753**	0.218	**0.147**
1990	**1.459**	3.475	**1126**	1783	**0.290**	0.459
1991	1.116	**0.288**	1004	**495**	0.345	**0.172**
1992	**0.534**	0.929	**577**	855	**0.260**	0.331
1993	**0.128**	0.223	**304**	425	**0.126**	0.184
1994	0.418	**0.374**	563	**533**	0.209	**0.188**
1995	0.429	**0.110**	610	**241**	0.207	**0.089**
(*b*) *Malaria*						
1989	**1738**	2717	**34**	42	**0.214**	0.253
1990	**598**	2215	**20**	41	**0.126**	0.266
1991	1226	**1079**	24	**23**	0.140	**0.138**
1992	3057	2544	*41*	*41*	0.388	0.401
1993	**2365**	2410	**36**	37	**0.263**	0.277
1994	2002	**1382**	39	**31**	0.367	**0.251**
1995	2701	**892**	49	**26**	0.532	**0.345**

*Boldface means the method ranks first; italics show similar behaviour.

A median test is used to evaluate the differences between the error measures of the two methods [17]. The chi-square values obtained for the test of comparing SARIMA and Bayesian models for each error criterion are non-significant at the 5 per cent level.

Figures 4 and 5 show the results of forecasting estimates of hepatitis A and malaria, for the seven periods and the two models.

For hepatitis A, Figure 4 shows that both methods have similar behaviour. For 1991, DLM follows more closely the behaviour of the corresponding observed values. The SARIMA model for the 1980.01–1991.13 period produced forecasts for 1992 that are closer to the observed data from month 1 to month 4, and underestimated forecasted values for the other months. For 1992, the Bayesian model had also a good performance for the first four months of 1992, but for later time periods the forecasts are usually much lower than the estimated values obtained with the SARIMA model. Overall, both models produced similar patterns for the forecasted values, having similar tendencies to overestimate or underestimate the observed values. These results suggest that, except for 1991 and 1992, the advantages of one approach over the other, based on the error measures, are only marginal.

Figure 5 shows that for malaria the forecasts produced for all periods are not different for the two models, and both cases are following closely the observed values most of the time. Again, here it is possible to see that the evaluation based solely in the error measures can be misleading since the pattern of forecasted values of both methods, even in cases where one has smaller performance errors, are very similar. An additional evaluation of the forecasting performance of both methods is obtained by plotting the forecast errors of the SARIMA method against the forecast errors of the DLM approach (Figures 6(*a*) and (*b*)). If both methods produce similar forecasts, all points would lie on a 45° line. In Figure 6, the points cluster close to this line showing that both methods are providing similar forecasts.

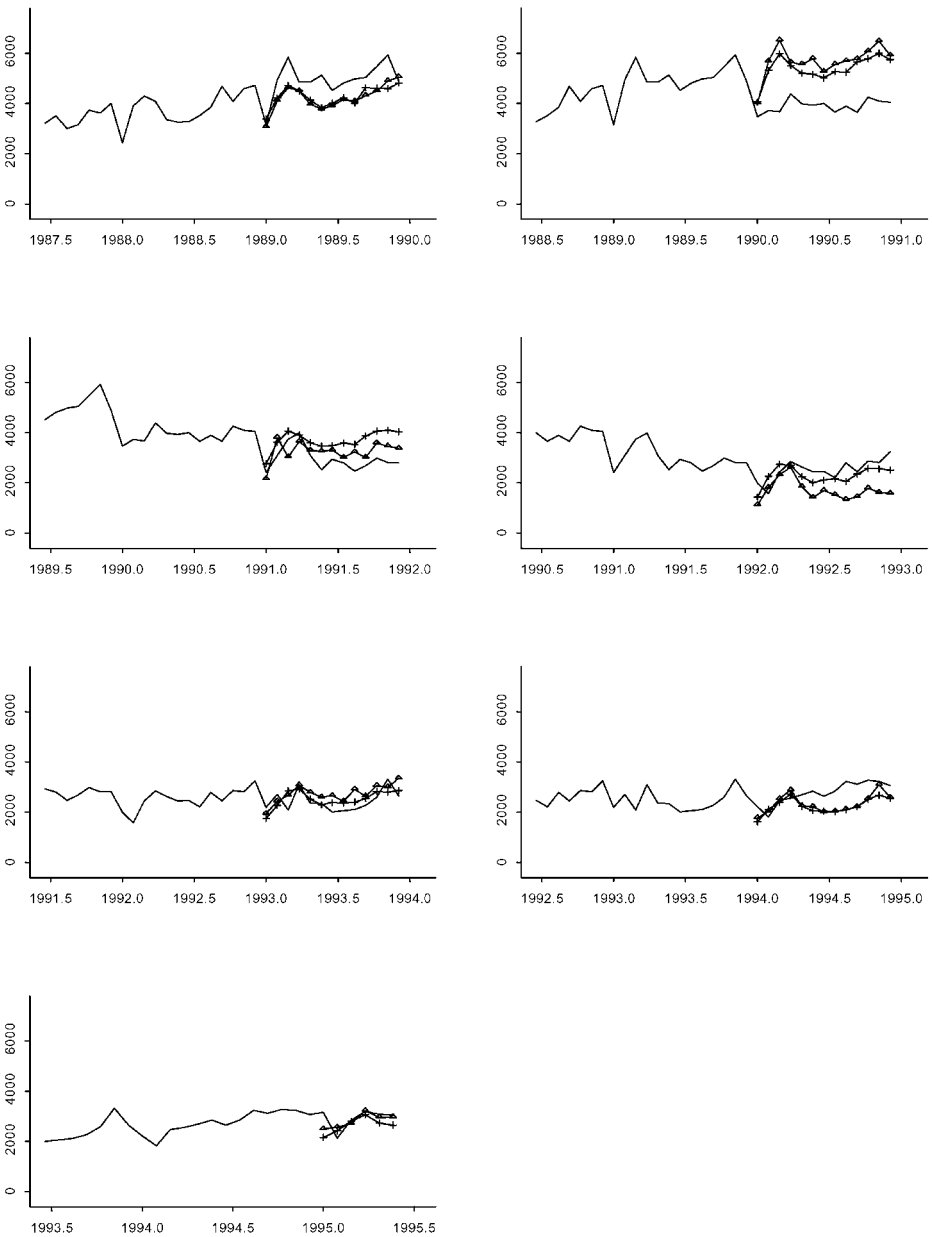

Figure 4. Forecasting estimates of hepatitis A for the seven periods: △, DLM forecasts; +, SARIMA. For display purposes observed number of cases are shown starting one and a half years before the forecasting period.

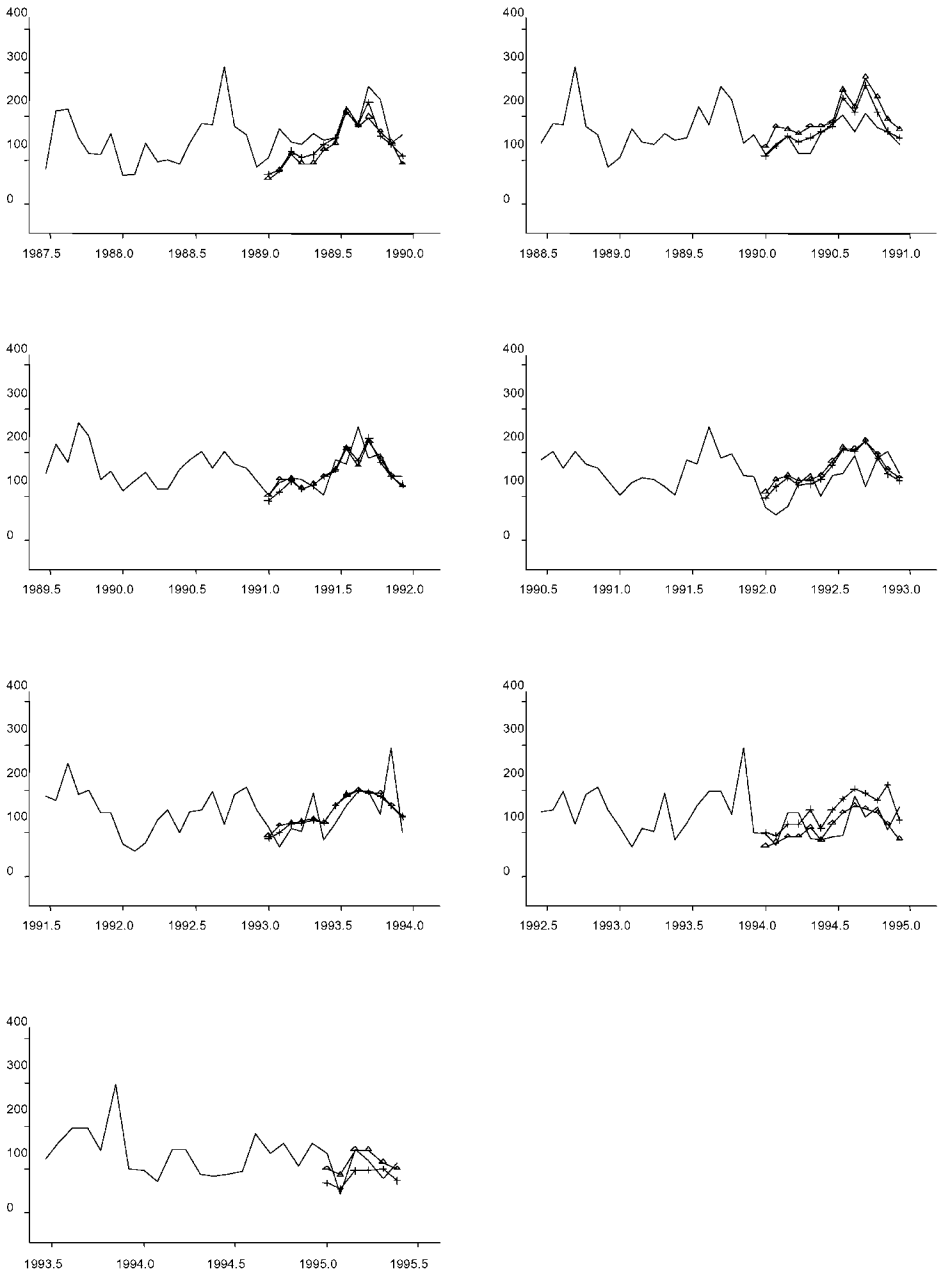

Figure 5. Forecasting estimates of malaria for the seven periods: △, DLM forecasts; +, SARIMA. For display purposes observed number of cases are shown starting one and a half years before the forecasting period.

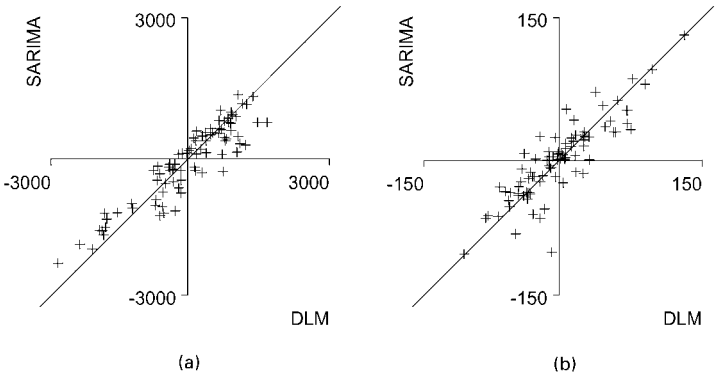

Figure 6. Comparison of the forecasting performance using the forecast errors: (*a*) for hepatitis A; (*b*) for malaria.

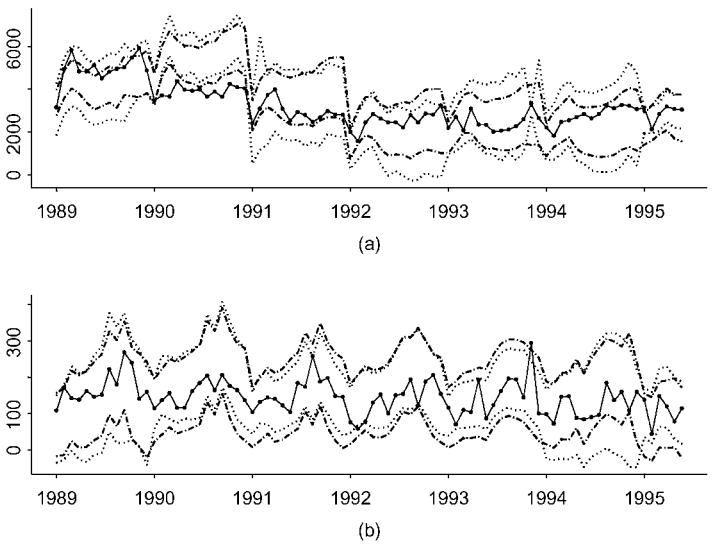

Figure 7. Estimated confidence intervals for the forecasts using DLM (···) or SARIMA (– · –): (*a*) for hepatitis A; (*b*) for malaria.

Since both methods can provide confidence intervals (CI) for the forecasts, a visualization of the CIs of the methods is shown in Figure 7. Analysis of these plots show that for hepatitis A, both methods fare badly for the period 1990.1 to 1990.13, producing overestimated forecasts and with lower CIs that are above the observed values.

5. DISCUSSION

Based on the three forecasting measured errors (MSE, MAE and MAPE), and the visualization of the forecasted values, forecasted errors and confidence intervals, the empirical evidence is that no one method dominates over the other. However, some tentative suggestions can be drawn to help the choice of a prediction method for use in public health surveillance.

The first method is the well-established SARIMA Box–Jenkins approach. The resulting models for both diseases have the same structures and the parameters do not change markedly for the seven forecasting time intervals. As in many epidemiological infectious disease time series, the data show seasonal pattern and trend components. The chosen SARIMA models fit the observations well; none of the series shows residuals that differ significantly from a zero mean white noise process. The accuracy of the forecast as measured by the three forecasting error criteria seems reasonable, particular for MAPE that has an average value of 0.236 for hepatitis A and 0.290 for malaria. One major problem with the Box–Jenkins approach is that the models, being based on the assumption of normality, are not entirely appropriate to count data such as health surveillance data. For more frequently occurring diseases the normality assumption may be less questionable. Here, the data used are suitable for modelling with this approach because they cover a long time period and the counts for the two diseases are large enough to treat them as continuous quantities. However, for diseases that are rare or surveillance data from regional or local level, which may have small counts and periods with no case occurrence, Box–Jenkins models may be inappropriate.

The other approach is the DLM method that has only recently been introduced in the forecasting literature. For hepatitis A and malaria, a seasonal dynamic linear model was selected as an appropriate model for both disease report series and all modelling time intervals. For hepatitis A, the discount factor for the trend decreased after 1991, suggesting more variation in its corresponding parameters. On the other hand, the discounts for seasonal parameters are all equal to unity, showing a constant value for them. Furthermore, the discounts for the precision are close to unity and, as a result, estimate of the observational variance, V, is based solely on the power, b, and the estimated expected value for μ. The monitoring procedure was an important addition to the modelling process since it allowed the selection of the model with best fit as each new observation arrives in the system. For malaria, it was needed only for monitoring the increasing of the level, whereas for hepatitis A, both upward and downward changes were detected. The DLM class of models is a more general modelling approach for count data. Since it does not require the normality assumption for the observations, it is more appropriate for forecasting applications involving small counts or with periods without case occurrence. However, the available softwares are not user friendly to allow the data and associated results to be more readily accessible and understood.

6. CONCLUSION

This investigation has applied two univariate forecasting approaches and compared their performance by using time series data belonging to the U.S. health sector. Hepatitis A and malaria data were chosen because they showed non-stationary features that seem to occur in many other epidemiological time series. Moreover, these time series fulfil the more restrictive conditions for applying the Box–Jenkins approach. To assess the forecasting abilities of the

two methods, standard approaches of each method have been used and three measures of the forecasting accuracy have been considered.

The use of time series models in epidemiology is increasing every day, and the consequences of their practical use are expected to improve the usefulness of public health surveillance systems. The comparisons, described in Section 4, show no significative difference between the two methods.

What recommendations can we make for the public health officials for the forecasting of the occurrence of diseases and other health events? The choice between the use of Box–Jenkins or DLM models depends on the intended use of the analysis. If one has epidemiological time series which can reasonably be assumed to be continuous-valued observations, a natural choice would be to apply the Box–Jenkins approach. This is so because the use of these models in many fields induced the development of new user-friendly software that simplifies and accelerates the modelling process.

If the research is analysing surveillance data from small geographical areas or related with health conditions that are rare, we recommend using the DLM approach discussed here or other methods developed for count data as already suggested by other authors [18]. Furthermore, the Bayesian nature of the DLM approach allows incorporation of subjective information, possibilities for forwarding interventions (for example, prospective changes in case definition) and the use of a direct associated monitoring scheme for detecting abnormal reported case occurrences.

Although Box–Jenkins and dynamic linear models had similar forecasting performances on our data, the latter is more flexible for dealing with different data scenarios, such as data sets from small areas or when new surveillance systems are established. The reported quantitative comparisons between SARIMA and DLM are strongly dependent of the time series and error measures chosen. Additional evaluation of both methods is needed, including using simulated data sets and comparison with other time series methodologies (for example, neural networks).

ACKNOWLEDGEMENTS

The authors are grateful to the referees for helpful comments. This work was partially supported by grants from MCT/FINEP/CNPq (Brazil). Thanks are also due to CAPES and FAPERJ for the scholarships for two of the authors.

REFERENCES

1. Thacker SB, Berkelman RL, Stroup DF. The science of public health surveillance. *Journal of Public Health Policy* 1989; **10**:187–203.
2. Stroup DF, Williamson GD, Herndon JL. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine* 1988; **8**:323–329.
3. Box EP, Jenkins GM. *Time Series Analysis Forecasting and Control.* Prentice-Hall: Englewood Cliffs, New Jersey, 1976.
4. Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962–1979. I. Time series forecasts of expected pneumonia and influenza deaths. *American Journal of Epidemiology* 1981; **113**:215–226.
5. Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962–1979. II. Percentage of pneumonia and influenza deaths as an indicator of influenza activity. *American Journal of Epidemiology* 1981; **113**: 227–235.
6. Stroup DF, Thacker SB, Herndon JL. Application of multiple time series analysis to the estimation of pneumonia and influenza mortality by age, 1962–1983. *Statistics in Medicine* 1988; **7**:1045–1059.
7. Helfenstein U. Box–Jenkins modelling of some viral infectious diseases. *Statistics in Medicine* 1986; **5**:37–47.

8. Waltier L, Richardson S. A time series construction of an alert threshold with application to S. Bovismorfbificans in France. *Statistics in Medicine* 1991; **10**:493–1509.
9. Williamson GD, Weatherby G. A monitoring system for detecting aberrations in public health surveillance reports. *Statistics in Medicine* 1999; **18**:3283–3298.
10. Brown RG. *Smoothing Forecasting and Prediction of Discrete Time Series*. Prentice-Hall: Englewood Cliffs, New Jersey, 1962.
11. Nobre FF, Stroup DF. A monitoring system to detect changes in public health surveillance data. *International Journal of Epidemiology* 1994; **23**(2):408–418.
12. Kalman RE. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering* 1960; **32**:35–45.
13. Stroup FD, Thacker SB. A Bayesian approach to the detection of aberrations in public health surveillance data. *Epidemiology* 1995; **4**:435–443.
14. West M, Harrison PJ. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag: New York, 1989.
15. Pole A, West M, Harrison J. *Applied Bayesian Forecasting and Time series Analysis.* Texts in Statistical Science. Chapman & Hall: New York, 1994.
16. Carter CK, Kohn R. On Gibbs sampling for state space models. *Biometrika* 1994; **81**:541–553.
17. Siegel S. *Nonparametric Statistics for the Behavioral Science*. McGraw-Hill: Tokyo, 1956.
18. Allard R. Use of time-series analysis in infectious disease surveillance. *Bulletin of the World Health Organization* 1998; **76**:327–333.

