STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:4202–4218
Published online 5 March 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2835

Automated time series forecasting for biosurveillance

Howard S. Burkom[1,*,†], Sean Patrick Murphy[1] and Galit Shmueli[2]

[1] *The Johns Hopkins University Applied Physics Laboratory, MD, U.S.A.*
[2] *Department of Decision & Information Technologies, Center for Health Information and Decision Systems, Robert H. Smith School of Business, University of Maryland College Park, MD, U.S.A.*

SUMMARY

For robust detection performance, traditional control chart monitoring for biosurveillance is based on input data free of trends, day-of-week effects, and other systematic behaviour. Time series forecasting methods may be used to remove this behaviour by subtracting forecasts from observations to form residuals for algorithmic input. We describe three forecast methods and compare their predictive accuracy on each of 16 authentic syndromic data streams. The methods are (1) a non-adaptive regression model using a long historical baseline, (2) an adaptive regression model with a shorter, sliding baseline, and (3) the Holt–Winters method for generalized exponential smoothing. Criteria for comparing the forecasts were the root-mean-square error, the median absolute per cent error (MedAPE), and the median absolute deviation. The median-based criteria showed best overall performance for the Holt–Winters method. The MedAPE measures over the 16 test series averaged 16.5, 11.6, and 9.7 for the non-adaptive regression, adaptive regression, and Holt–Winters methods, respectively. The non-adaptive regression forecasts were degraded by changes in the data behaviour in the fixed baseline period used to compute model coefficients. The mean-based criterion was less conclusive because of the effects of poor forecasts on a small number of calendar holidays. The Holt–Winters method was also most effective at removing serial autocorrelation, with most 1-day-lag autocorrelation coefficients below 0.15. The forecast methods were compared without tuning them to the behaviour of individual series. We achieved improved predictions with such tuning of the Holt–Winters method, but practical use of such improvements for routine surveillance will require reliable data classification methods. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: forecasting; regression; biosurveillance; exponential smoothing; time series; preconditioning

*Correspondence to: Howard S. Burkom, The Johns Hopkins University Applied Physics Laboratory, MD, U.S.A.
†E-mail: howard.burkom@jhuapl.edu

Contract/grant sponsor: Office of State and Local Government Coordination and Preparedness (SLGCP); contract/grant number: #8AUAS5
Contract/grant sponsor: United States Department of Homeland Security

Received 17 April 2006
Accepted 20 December 2006

1. INTRODUCTION

This paper examines and compares methods for the automatic preconditioning of health indicator data to enable the timely prospective monitoring required for effective syndromic surveillance. The objective of syndromic surveillance is to complement traditional public health monitoring with routine, automated data analysis. For this purpose, both clinical and non-clinical data sets may be used, including records of hospital emergency department visits, physician office visits, outpatient clinic visits, over-the-counter (OTC) remedy sales, school and work absenteeism rates, and others. Automated data collection and analysis systems have the potential to widen the scope of traditional surveillance while providing a safety net for missed outbreak indications. Before the year 2000, prospective health surveillance dealt mainly with weekly or monthly data [1, 2], but since the terrorist attacks of September 2001, attention has shifted to alerting based on daily and faster data rates. The urgency caused by the perception of a bioterrorism threat has led to the rapid development and activation of surveillance systems that often employ technologies developed in other fields.

The statistical process control (SPC) community has developed a wealth of robust, sensitive monitoring methods in the form of control charts and related methods [3]. Although such charts have been implemented for a wide variety of health-monitoring purposes [4], some implementations monitor data that violate basic assumptions required by these methods [5] yielding alerting methods with uncertain detection performance. This problem highlights an inherent obstacle to the use of traditional SPC methods for syndromic surveillance: the nature of the data. Syndromic data streams are based not on physical science, as are manufacturing processes, but on changing population behaviour and evolving data acquisition and classification procedures. Reliable detection performance requires methods that can adapt to the resulting changes in background behaviour.

To further complicate the health-monitoring problem, data are aggregated at various levels by time, location, and diagnostic specificity according to epidemiological objectives and jurisdictional constraints. From the detection point of view, these aggregation decisions determine background noise levels that must be understood to achieve high sensitivity at manageable alert rates. Aggregation decisions may result in characteristic time series features such as day-of-week effects and seasonal patterns [6]. Numerous epidemiologists have requested an automated capability to customize case definitions or to adjust them as a situation develops [7], and such flexibility requires rapid selection of appropriate analysis methods. Similar surveillance requirements have caused a surge in research by statisticians, informaticists and researchers in related disciplines towards the development of more advanced monitoring algorithms. While this research has produced a variety of new algorithmic approaches [8], the utility of these new techniques for syndromic surveillance remains to be determined.

An alternative approach to developing more advanced monitoring algorithms is to precondition the data to derive residual data streams appropriate for traditional monitoring tools. The derivation of residuals free of temporal dependence is crucial for routine robust detection performance and has a long history in the SPC community [9]. Indeed, several authors have modelled biosurveillance data to obtain such residuals [10, 11]. For seasonal forecasting, a primary obstacle for these methods is the difficulty of adjusting to short-term trends and year-to-year seasonal variations, and the benefits of adjusting are illustrated in the results of this paper. Given the growing emphasis on timely alerting using daily and faster data acquisition rates, it is essential that the alerting methods

DOI: 10.1002/sim

of syndromic surveillance manage this obstacle. Management approaches include:

(a) non-adaptive regression modelling with sufficiently frequent expert review and adjustment [11];
(b) adaptive regression modelling with a sliding baseline to capture recent data behaviour [12];
(c) inclusion of local variations in modelling, as in ARIMA methods [13];
(d) generalized exponential smoothing to explicitly modify forecasts by changes in expected modes of recent behaviour [14].

Approaches (a) and (b) may be seen as engineering solutions attempting to make the global approach robust. Their success depends upon ongoing data analysis and the consistent applicability of the chosen model. Approach (c) is theoretically appealing because a local error process is directly modelled, though the choice of appropriate ARIMA parameters may require sophisticated analysis as well as periodic review [15]. Approach (d) is not a global behaviour model [8] but a recursive forecasting method notable for its simplicity and comparative performance to more expensive methods. Comparing (c) and (d), Chatfield remarked, 'Practical considerations rule out... [the ARIMA approach]...if there are insufficient observations or ...expertise available' [14]. In view of the set of non-stationary syndromic data streams to be modelled, sometimes with no data history or opportunity for prior analysis, we do not test an ARIMA approach in this study.

One of the motivations for this study was operational experience since 1999 with several versions of the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) [16]. Such systems typically do not have years of prior data to examine, and when they do, historical data are often unrepresentative of current behaviour for a variety of reasons: changes in treatments; changes in coding; changes in population behaviour; evolving informatics—faster data rates; changes in reporting practices, etc. In the data streams characteristic of syndromic surveillance and typically encountered in ESSENCE biosurveillance systems steady-state behaviour is the exception rather than the rule. Without the luxury of studying large data sets retrospectively, optimal modelling is not practical, and methods are required to fit the constraints of their application environment. Furthermore, the growing set of disparate evidence available at increasing acquisition rates, combined with increasing concerns over bioterrorism and naturally occurring disease such as new influenza strains, make automated analyses both necessary and challenging and forms the practical motivation for this paper.

The paper is organized as follows: Section 2 presents our three preconditioning methods; Section 3 describes the data and study design; Section 4 examines the results and compares the performance across methods, and Section 5 is a summary discussion.

2. FORECAST METHODS

In this section we describe the three time series forecasting methods that were compared in this study.

2.1. Non-adaptive regression model

The first forecast method follows Brillman *et al*. [11], fitting an ordinary least-squares, loglinear model to a set of training series values to obtain regression coefficients. These coefficients are then used to forecast values beyond the training data without adjusting for subsequent changes in time-series behaviour.

Setting Y_t as the number of syndromic hospital visits on day t, the authors model the 'started log' $Y_t^* = \log(Y_t + 1)$ as

$$Y_t^* \sim \left[\sum_{i=1}^{7} c_i I_{i,t}\right] + [c_8 + c_9 \times t] + [c_{10} \times \cos(kt) + c_{11} \times \sin(kt)]$$

where c_1–c_7 are coefficients for day-of-week effect indicators, c_8 is a constant intercept, c_9 is the slope of a long-term trend, and c_{10}–c_{11} are coefficients for continuous harmonic terms representing seasonal trends, with $k = 2\pi/365.25$ to give a 1-year sinusoidal period for these terms.

The reason for transforming the original daily counts into log scale is to capture the multiplicative nature of the effects of the trend and seasonal components. In other words, counts on different days of the week differ on a percentage scale rather than by fixed amounts. Additionally, the distribution of daily counts tends to be right skewed rather than normal (as assumed by linear regression models). The log transformed data, however, are usually much closer to a normal distribution. The addition of 1 assures that the argument of the logarithm is positive.

In the study below, we added one covariate, a holiday indicator, I_t^{hol}, to reduce the distortion of predictably reduced holiday observations on the estimation of the other regression coefficients. The indicator was assigned the value $I_t^{\text{hol}} = 1$ on the following 10 holidays: New Year's Day; New Year's Eve; Christmas; 4th July; Thanksgiving; Memorial Day; Labor Day; Veteran's Day; Martin Luther King Jr. Day; and Columbus Day. On all remaining days, $I_t^{\text{hol}} = 0$.

Application of this non-adaptive model for routine surveillance makes two underlying assumptions. The first is that the chosen covariates—here the day-of-week, trend, and seasonal terms—and their model representations are sufficient to capture the systematic background behaviour of the time series, i.e. the behaviour that is not relevant to an outbreak signal of concern. Substantial data history is usually required for seasonal modelling using harmonic terms; Brillman *et al*. used over 8 years of training data. The second assumption is that the relationship between the syndromic counts and the covariates does not change from the training data and can be used for out-of-sample prediction. Those authors present results to support the former assumption and discuss practical means to correct for the failure of the latter. The next method was devised to adapt this regression approach to the changing data behaviours and limited histories often available in automated health surveillance.

2.2. Adaptive regression model

The second forecast method is an adaptive regression model with a sliding 8-week baseline interval [12]. This model is similar in form to the global one given above, again using $Y_t^* = \log(Y_t + 1)$:

$$Y_t^* \sim \left[\sum_{i=1}^{7} c_i I_{i,t}\right] + [c_8 + c_9 \times t] + [c_{10} \times I_t^{\text{hol}}]$$

where c_1–c_7 are coefficients for day-of-week indicators, c_8 is a constant intercept, c_9 is the slope of a linear trend, using a centred ramp function, and c_{10} is a coefficient for a holiday indicator, as defined earlier.

This method recomputes the regression coefficients for each forecast using only the series values from the eight weeks before the forecast day. The short baseline is intended to capture

DOI: 10.1002/sim

recent seasonal and trend patterns. The sinusoidal covariates cannot be used because the baseline interval is a small portion of their one-year period that would cause the harmonic terms to be nearly linear over some baseline intervals. The holiday indicator was added to avoid exaggerated forecasts on known holidays and to avoid the computation of spurious values for c_1–c_7 when holidays occurred in the short baseline interval. A similar model is applied for anomaly detection in ESSENCE biosurveillance systems when an automated goodness-of-fit criterion is satisfied. In those operational implementations, a post-holiday indicator is also added to account for increases following a holiday. In addition, the two days of data immediately before the forecast day are omitted from the baseline used to estimate the regression coefficients. This two-day buffer is imposed to reduce the chance of contamination of the baseline with the early portion of a gradual outbreak signal and to avoid the resulting loss of sensitivity.

2.3. Generalized exponential smoothing

Our third forecast method was generalized exponential smoothing using the multiplicative Holt–Winters approach. This method is not a regression or other data model but a data-driven smoother that can account for systematic series behaviour such as trends and cyclic effects. Neither older [4] nor more recent [5] survey articles on health surveillance methodology study the use of this method, though Shmueli and Fienberg [8] cite its popularity in business and advocate its application to health surveillance. We selected it for comparison to the above models because of its simple formulation and straightforward adaptation to local behaviour changes. The following paragraphs introduce this approach and discuss adaptations for daily health data monitoring.

In their simplest form, exponential smoothing forecasts are based on a weighted average of past observations, where the weights decay exponentially with the age of the observations. Forecasting with simple exponential smoothing assumes that the underlying series comprises two components: a time-varying 'level' L_t and random noise. We use the data to update our estimate of the level. Consider a time series $Y_1, Y_2, Y_3, \ldots$. The updating equation for L_t is

$$L_t = \alpha Y_t + (1-\alpha) L_{t-1} \tag{1}$$

for a fixed smoothing coefficient α, $0<\alpha<1$. The current level is thus a weighted average of the previous level and the new observation. Rewriting this equation as

$$L_t = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots \tag{2}$$

demonstrates the exponential decay of the weights. The k-step ahead forecast, given information until time t, is the most recent level estimate:

$$\hat{Y}_{t+k} = L_t \tag{3}$$

This smoothing may be seen as the expression of a natural learning process in which observations have decreasing influence as they age, and the most recent estimate, e.g. of a sample mean, is modified in fixed, linear fashion by the most current observation [17].

This procedure may be used to generate forecasts, as in this paper, or to form the test statistic for the EWMA control chart [3], which has been adapted for biosurveillance [12, 18, 19]. Note that an initial value L_1 is required for the recursion. A mean estimate for Y_t is preferred for this initial value, but in its absence the first data value is often used, with negligible effect on smoothed values beyond a short warmup interval depending on the smoothing constant α. The variance

DOI: 10.1002/sim

of L_t beyond the warmup interval is

$$\sigma_L = (\alpha/(2-\alpha))\sigma_Y \tag{4}$$

where σ_Y is the variance of the original series. In practice, a range of values for α is usable, and lower values are preferred to reduce this variance; values between 0.1 and 0.3 are often chosen.

Exponential smoothing based on the data level alone should not be used for forecasting series that exhibit seasonality or trend [20, p. 730]. Similarly, control charts such as Xbar, EWMA, and CUSUM should not be applied directly to time series with cyclic or other systematic behaviour because of the resulting bias. For example, if visit counts on Monday are generally high because clinics reopen after the weekend, then chart-based methods that ignore this feature will be biased towards alerting on Mondays. If the systematic behaviour is consistent, regression may be applied to derive residuals appropriate for these charts [9]. Another approach is to generalize the exponential smoothing concept to express changing trends and cyclic effects as well as the changing level. In industrial and financial time series forecasting, a well-known implementation of this generalization is the Holt–Winters forecasting method [21, 22]. Along with the level L_t, the method includes two additional recursive terms, one for the trend T_t and one for a seasonal component S_t. The k-step ahead forecast is given by (see, e.g. [23])

$$\hat{Y}_{t+k} = (L_t + kT_t)S_{t+k-M} \tag{5}$$

where M is the number of seasons in a cycle (e.g. for monthly data with a calendar-month effect, we would choose $M = 12$) and L_t, T_t, and S_t are updated as follows:

$$L_t = \alpha\frac{Y_t}{S_{t-M}} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{6}$$

$$S_t = \gamma\frac{Y_t}{L_t} + (1-\gamma)S_{t-M} \tag{7}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{8}$$

The three updating equations tell us that the updated component at time t is a weighted average of the (adjusted) previous estimate and the most recent information that arrives at time t. For example, in equation (6) we see that the level is updated by taking the most recent level estimate (L_{t-1}), adjusting it for trend (T_{t-1}), and averaging it with the most recent seasonally adjusted observation (Y_t/S_{t-M}). Similarly, the seasonal component (equation (7)) is updated by averaging the previous seasonal component (S_{t-M}) with the most recent de-trended observation (Y_t/L_t). Finally, the trend is updated by averaging the previous trend component (T_{t-1}) with the difference between the two most recent level estimates (the trend is defined as the change in level).

We use the multiplicative version of the forecast in (5) and updates (6) and (7) because the cyclic effects in our syndromic time series are generally proportional to the level L_t. In other words, these effects act as daily count multipliers rather than by adding fixed amounts. An additive formulation is also available [14].

This method requires three smoothing constants, α, β, and γ. Practical issues arising in generalized exponential smoothing were discussed in detail in [15], and we summarize some of those findings that are relevant to the forecasting of syndromic time series. First, the chosen updating equations should be the simplest formulation that can capture the expected data features. Second, in the Holt–Winters formulation, forecasting beginning with $\hat{Y}_s$ requires initial values for the level,

DOI: 10.1002/sim

trend, and the M seasonal components. The choice of these initial values is more important than in simple exponential smoothing, especially for the cyclic behaviour terms. If the cyclic behaviour is consistent and representative starting values for $c_0, c_1, \ldots, c_{M-1}$ are available, then a γ value near zero may give optimal forecasts. If this behaviour changes seasonally or changes fairly often as a result of artefacts in the data acquisition chain, then a larger γ value is indicated, and analogous simpler guidance applies to the choice of α and β. Third, the search for optimal values of the smoothing coefficients is non-trivial and does depend on the initial values. A common approach is to find values of α, β, and γ that minimize the mean or median absolute error, mean or median absolute per cent error (MedAPE), or a similar measure [15]. However, the solution surface is not convex, and gradient-type searches generally will not find the global optimum without a good starting guess, so a grid search should be used. We employed a coarse grid search by computing the median absolute deviation (MAD) for all combinations of smoothing constants in $\{0, 0.1, 0.2, \ldots, 1\}$ and selecting the combination that minimized this measure. From our experience, a finer search is likely to overfit the data.

We implemented the Holt–Winters method according to the properties of the time series chosen for this study, which are typical of daily time series from moderate to large geographic areas for clinic visits, physician office visits, OTC sales and ED visits from some hospitals. Our cyclic component included $M=7$ day seasons to capture the common day-of-week effects. Note that this method, like the regression method using day-of-week indicators, does not assume a particular weekly pattern but can be used for time series where weekend counts drop slightly or nearly vanish, where Saturday counts are elevated as in some OTC sales data, or where the pattern is more unusual, as in visits of clinics with fixed weekly schedules. With a positive γ value, this smoothing method can also adapt to changes in this pattern.

To avoid forecasts of negative syndromic counts, we set zero as a lower bound for the recursively computed level L_t. We avoided zero divisors in the updating equations for L_t and S_t by setting

$$L_t = L_{t-1} + T_{t-1} \quad \text{if } S_{t-M} = 0 \tag{9}$$

and

$$S_t = S_{t-M} \quad \text{if } L_t = 0 \tag{10}$$

Another modification was to suppress the updating of the level and trend when the forecast error was very large, to avoid incorrect learning from outliers. In other words, updating equations (6) and (8) were not applied when forecast errors were very large. In such cases, the level and trend from the previous time point were retained. For the data of this study, we applied this condition by updating only when the absolute forecast error was no more than half the forecast value, and we chose this ratio criterion empirically. For time series with small counts, a different criterion should be applied.

3. STUDY DESIGN

3.1. Study data

The goal of our study was to compare the predictive performance of these forecast methods to obtain residuals free of bias-causing systematic behaviour for input to standard alerting algorithms. The study data were time series of aggregated, de-identified counts of health indicators derived from the BioALIRT program conducted by the U.S. Defense Advanced Research Projects Agency

DOI: 10.1002/sim

(DARPA) [24]. Formal agreements to use these data were signed by the authors, and others wishing access may contact the corresponding author for the required procedures.

Three types of daily syndromic counts were represented: military clinic visit diagnoses, filled military prescriptions, and civilian physician office visits. The BioALIRT program categorized the records from each data type as respiratory (RESP), gastrointestinal (GI), or other, and the data were gathered from 10 U.S. metropolitan areas with substantial representation of each data type. This study used the RESP and GI data for all three data types from five of the cities for a total of 30 time series, each including syndromic counts for 700 days. The RESP time series showed consistent day-of-week effects and seasonal trends, while the GI time series showed only day-of-week effects, as seen in Figure 1. To restrict attention to authentic data representing routine consumer behaviour and disease trends but which do not contain data quality problems such as changing participation of data providers, we excluded time series in which temporary dropouts and permanent step increases were evident. These exclusions allow the comparison of forecasts of data based only on consumer behaviour, not on artefacts in the data chain. While an operational health-monitoring system must manage such data problems, the goal of this investigation was to isolate the issue of removing systematic data behaviour from these problems and from the choice of control charts and other alerting methods that use the data residuals as input. The remaining data included 10 time series of RESP counts and six time series of GI counts. All three forecast methods were applied to these 16 time series.

The first half of each time series, consisting of 350 daily syndromic counts, was used as training data. The second half of each series was used as test data for comparing predictions. For the non-adaptive regression forecasts, the entire 50 weeks of training data were used for model fitting. This training set is well short of the 8+ years used in [11] but does include nearly a full period of the sinusoidal covariates. Moreover, multiyear data sets, especially with consistent background distributions, are often unavailable in practical surveillance. The adaptive regression method used the 8-week sliding window for model fitting, so each forecast in the test interval was based on inference using only the preceding 56 days of data. Thus, most of the interval used to train the global model was not used by the adaptive regression. For the third method, Holt–Winters recursion

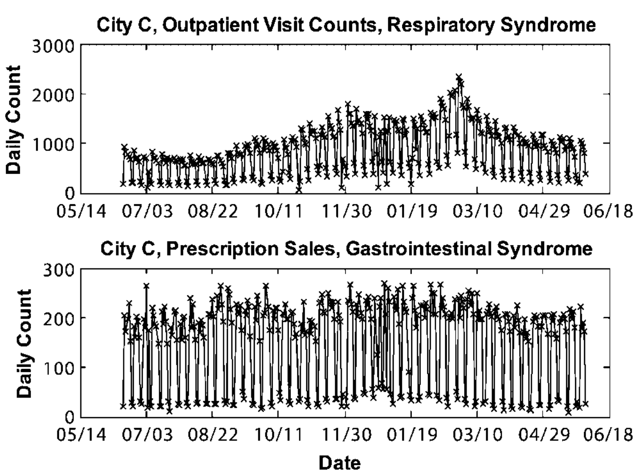

Figure 1. Sample time series of daily counts from respiratory and gastrointestinal syndrome groups.

Statist. Med. 2007; **26**:4202–4218
DOI: 10.1002/sim

was performed on all 700 days of each time series. However, due to the exponential weighting, most of the training data had negligible effect on the test interval forecasts.

3.2. Forecast method implementation

In his comparative discussion of Holt–Winters and Box–Jenkins forecasting, Chatfield noted [14] the importance of comparing methods that are either both automatic or both non-automatic. The term 'automatic' referred to the use of a preselected ARIMA model for the Box–Jenkins approach and a preselected set of smoothing coefficients for the Holt–Winters method. Given the limited data histories and multiple, sometimes improvised syndrome classifications used in biosurveillance, we restricted our regression models and Holt–Winters smoothing formulation to automatic methods. We closely followed the models described in 2.1 and 2.2 for non-adaptive and adaptive regression.

For the Holt–Winters implementation, smoothing coefficients and starting values were required. We chose sets of smoothing coefficients with an approximate clustering approach based on grid search results from all of the time series. For the RESP series, the chosen coefficients were $\alpha=0.4$, $\beta=0.0$, $\gamma=0.15$, and for the GI series we used $\alpha=0.1$, $\beta=0.0$, $\gamma=0.15$. In a 'non-automatic' forecasting experiment noted in the results section, we did obtain improved predictions by optimizing coefficients for a single time series, as in [14], but the applicability of such optimization for future forecasts and across different time series is unclear, and thus overfitting must be avoided. For starting values, we set the level L_1 in each series to the mean of the first 28 days of data. The trend term T_1 was set to zero and not updated in equation (8) because β was also zero. Our experiments suggested that trends in these series types were local in nature, so that the level L_t updates in equation (6) could account for these trends as well as explicit T_t terms. The initial day-of-week factors $S_1, \ldots, S_7$ were set to 1 so that all of the cyclic adjustment would be 'learned' by updating equation (7). For smoothing to obtain more prompt forecast capability without the 350-day warmup period of our study, we found the initialization recommendations of [15] readily adaptable.

3.3. Performance measures

We compared the predictive accuracy of our forecast methods using the residuals from the 350-day test interval. For each method, these residuals were computed as the observed value minus the forecast for each day. We applied several measures to each set of residuals. The root-mean-squared error (RMSE), or square root of the mean of the squared residuals, is a commonly used measure that gives weight to all error magnitudes, including those from the forecasts that are poor because of known effects such as holidays. Thus, a single large outlier may influence this measure significantly. For a more robust statistic, we also computed the MAD, a broader central tendency measure that ignores the magnitude of the worst (and best) errors. Our third measure was the MedAPE, which gives an idea of the typical percentage error and also allows comparisons across different series.

4. STUDY RESULTS

4.1. Comparison of residuals for 1-day ahead predictions

We first look at general features of the residuals from the forecast methods. Plate 1(a), a plot representative of the results, uses dark dots to show the counts of daily clinic visits classified in the respiratory syndrome over the 350-day test interval from one of the time series chosen for the

DOI: 10.1002/sim

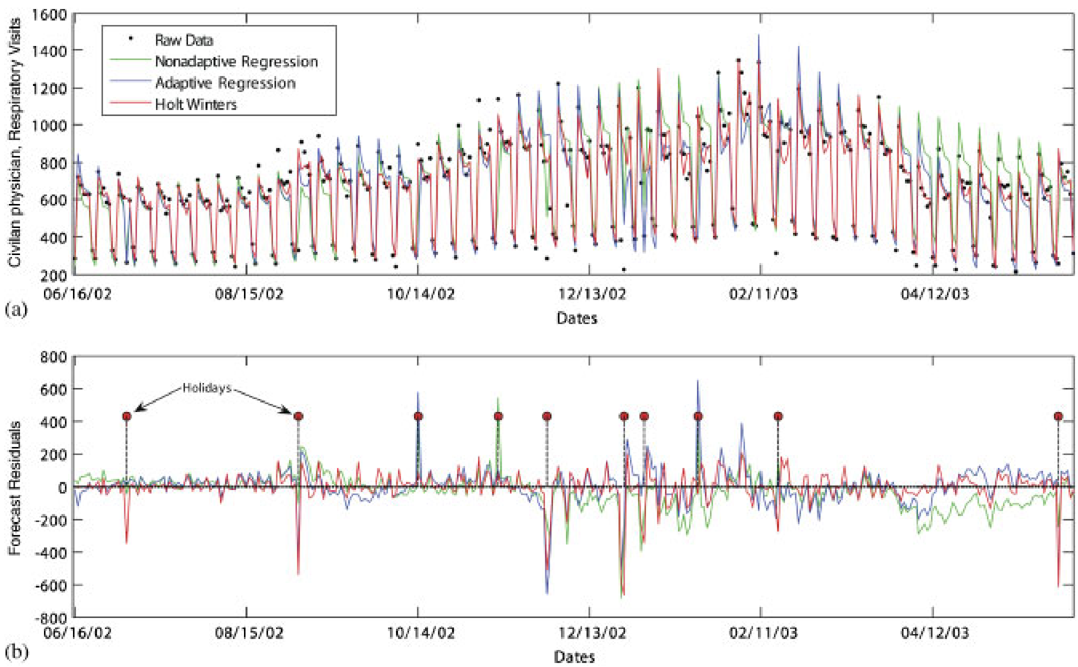

Plate 1. Respiratory syndrome time series with forecasts and residual comparisons with 10 calendar holidays indicated: (a) civilian physician, respiratory visits for City B with 1-day ahead forecasts and (b) forecasts residuals.

DOI: 10.002/sim

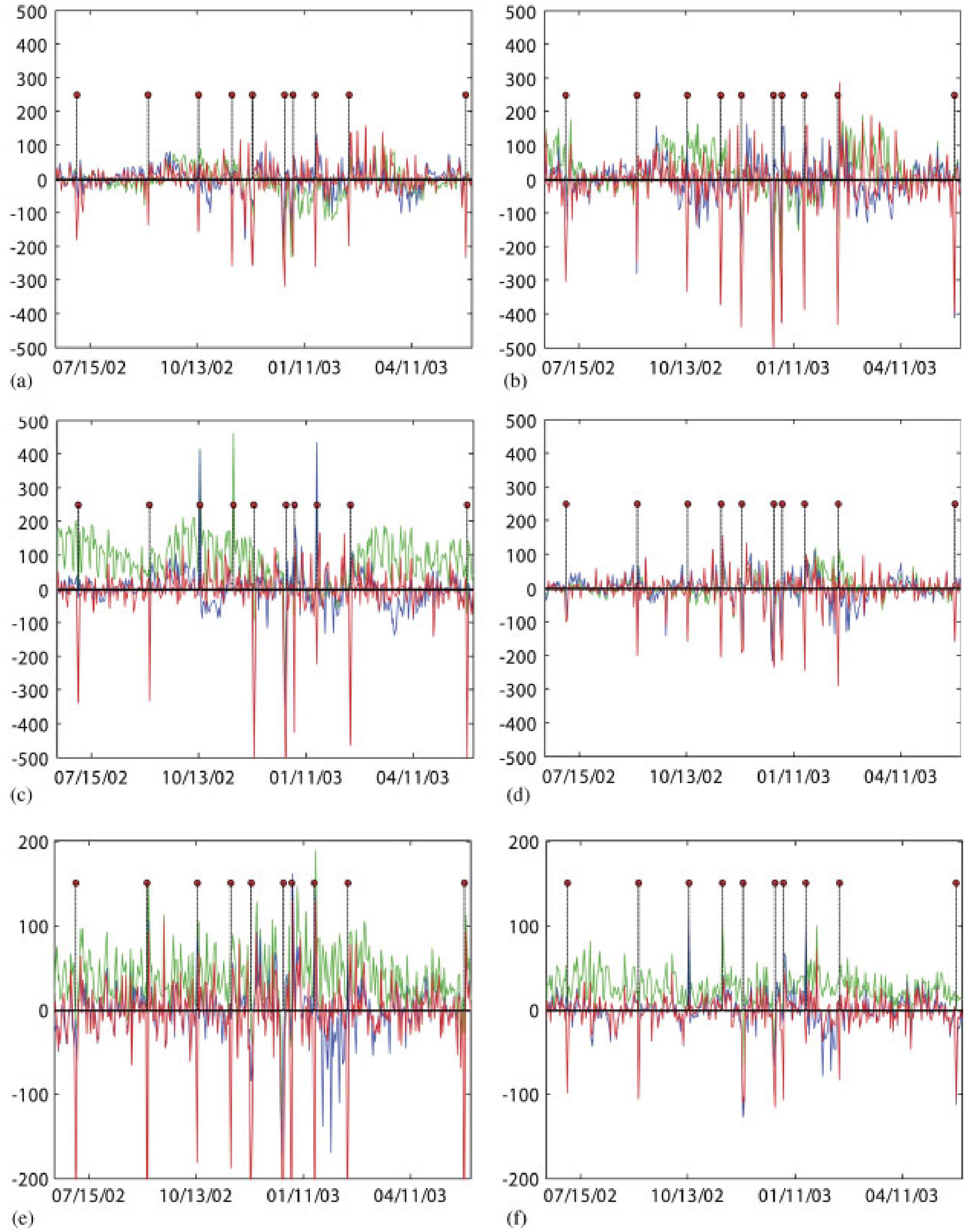

Plate 2. Forecast residual comparison plots for (a–d) respiratory syndrome data and (e and f) gastrointestinal syndrome data. Residuals from non-adaptive regression are in green, from adaptive regression in blue, and from Holt–Winters smoothing in red: (a) City A: military clinic visits; (b) City A: military prescriptions; (c) City A: civilian physician visits; (d) City D: military clinic visits; (e) City A: military prescriptions; and (f) City A: civilian physician visits.

study. On the same axes, the green, blue, and red curves show forecasts made with non-adaptive regression, adaptive regression, and Holt–Winters methods, respectively. Note the day-of-week effects and the short-term trends that appear in the observed daily counts. The global regression forecasts have no capability to adapt to the short trends, while the adaptive regression forecasts can adapt to trends on the order of the 8-week baseline. The generalized exponential smoothing is more suited to local linear behaviour. Non-adaptive regression forecasts may be very accurate if the seasonal behaviour in the test interval is similar to the behaviour in the training interval. However, if a season is unusual relative to the training period, then the non-adaptive method may overpredict or underpredict for the entire season; note the consistent underprediction indicated by the positive residuals in the spring of 2003.

Plate 1(b) is a companion plot showing the corresponding residuals for the three forecast methods with the same colour coding; i.e. the curve for each method is the difference of the observed counts and the forecasts. The red circles indicate 10 calendar holidays on which observed counts are reduced; note the overpredictions of the adaptive regression and Holt–Winters methods on these days. Our Holt–Winters implementation avoided recursive updates but did not avoid forecast errors on such outlier days, while the 8-week baseline of the adaptive regression sometimes did not include enough holidays to reduce forecasts appropriately. Both adaptive methods need special treatment for rare but expected model exceptions.

Plate 2 is a group of comparison plots analogous to Plate 1(b) for the other nine RESP series and for the six GI series of the study. Note the sustained underprediction of the non-adaptive regression in several of the plots. The Holt–Winters and adaptive regression forecasts are generally comparable, though there are intervals such as February 2003 in RESP series (e) where the short, recent adaptive regression baseline can be misleading. In applying the non-adaptive regression method to RESP series (i) and possibly to some others, the mismatch of the data in the training and test intervals would soon be evident in a prospective application.

Table I gives a quantitative comparison of the 1-day-ahead forecasts using the MAD, MedAPE, and RMSE measures. Forecast comparisons depend on the measure chosen. Judging from the medians, which ignore the sizes of errors on all unusual days, the Holt–Winters forecasts are almost uniformly the best among the methods. Use of the medians drops the effects of the 10 calendar holidays as well as the post-holiday effects from the 350-day test interval. By contrast, in the mean calculations of the RMSE criterion, the influence of the extreme errors on holidays is retained. Thus, if we compare residuals with the RMSE criterion, the inclusion of all 10 holidays worsens the Holt–Winters forecasts for most of the time series relative to the other methods, and the adaptive regression forecasts look best. The MedAPE measures, comparable in scale over the 16 test series, averaged 16.5, 11.6, and 9.7 for the non-adaptive regression, adaptive regression, and Holt–Winters methods, respectively.

Plate 3 is a typical view of the residuals stratified by day-of-week and by season. Residuals by season were compared using cutoff dates of 22 March–21 June for spring, 22 June–21 September for summer, 22 September–21 December for fall, and 22 December–21 March for winter. The plot shows the MAD criterion to remove holiday effects. The chosen time series is the first RESP series, summarized in the first column at the top of Table I. The overall Holt–Winters MAD value is lowest for each season, and this advantage is also seen for all days except Mondays. Similar analyses for some of the other series also showed increased prediction errors on Mondays for the exponential smoothing, and a more careful choice of smoothing coefficients might reduce this problem. Despite the seasonal harmonic terms in the non-adaptive regression, the residual MAD is most strongly a function of season for that method.

DOI: 10.1002/sim

Table I. Comparison of 1-day-ahead forecasts for 10 respiratory count series and six gastrointestinal count series, using median absolute residual, median absolute per cent error, and root-mean-square residual.

	Respiratory count series	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
	Non-adaptive regression	22.6	31.7	83.1	**13.3**	27.3	35.6	43.9	53.7	76.7	131.1
MAD	Adaptive regression	20.0	29.3	25.8	16.3	27.7	22.9	28.0	47.6	68.1	39.0
	Holt–Winters	**17.5**	**26.7**	**21.9**	**13.3**	**26.3**	**19.2**	**24.7**	**30.6**	**50.7**	**33.6**
	Non-adaptive regression	11.7	11.6	18.8	**10.0**	10.5	19.1	19.8	8.6	10.4	32.6
MedAPE	Adaptive regression	9.7	9.8	6.5	12.0	12.4	14.0	14.1	7.8	10.4	9.8
	Holt–Winters	**9.1**	**9.3**	**5.3**	10.4	**9.9**	**11.3**	**12.2**	**5.1**	**7.7**	**8.1**
	Non-adaptive regression	44.8	69.3	110.0	**34.3**	**71.2**	74.6	65.1	113.3	**197.9**	244.1
RMSE	Adaptive regression	**41.1**	**61.9**	**71.6**	41.6	87.3	57.9	**54.2**	112.3	212.6	121.7
	Holt–Winters	54.7	87.5	81.4	49.2	98.6	**52.9**	65.1	**100.1**	201.5	**92.6**

	GI count series	(a)	(b)	(c)	(d)	(e)	(f)
	Non-adaptive regression	13.5	35.6	24.3	38.9	24.1	21.1
MAD	Adaptive regression	9.2	15.9	9.1	25.3	11.6	13.5
	Holt–Winters	**8.7**	**15.7**	**8.0**	**24.6**	**10.2**	**12.2**
	Non-adaptive regression	15.3	17.8	23.4	10.7	22.6	20.5
MedAPE	Adaptive regression	**10.4**	**8.2**	9.3	7.3	10.8	13.4
	Holt–Winters	10.6	8.3	**8.7**	**7.1**	**8.9**	**12.1**
	Non-adaptive regression	19.7	51.5	32.2	**58.3**	43.5	32.9
RMSE	Adaptive regression	**17.0**	**42.0**	22.5	59.6	34.9	**27.3**
	Holt–Winters	19.1	54.8	**21.6**	76.5	**34.1**	28.6

Note: Bold font signifies the best forecast among the three methods according to each residual measure for the corresponding time series.

4.2. Comparison of residual autocorrelation properties

As noted in Section 1, the motivation for this study is to obtain residuals suitable for input to control charts and similar anomaly detection algorithms. Most of these algorithms have the underlying assumption that the input observations are independent. Thus, the degree of serial correlation among residuals is a relevant criterion for these forecast methods. For the three tested methods, we computed the autocorrelation functions of the residuals from the forecasts of each of the 16 chosen data series. Plates 4(a) and (b) show representative plots for a RESP series with both day-of-week and seasonal variation and for a GI series with only day-of-week features. The top plots show the respective data series, and the middle plots show the data correlation coefficients for lags of 0–28 days. The distinct day-of-week effects are evident in the coefficient values for lags that are multiples of 7. For all 16 time series used in the study, the 7-day lag coefficient exceeded 0.95.

The bottom plots show the correlation coefficients for the forecast residuals, with the same colour coding as in previous plates. For all plotted lag values, the non-adaptive method residuals show the most correlation, though a review of Table I shows that the chosen time series gave the best relative forecasts for this method. The autocorrelation coefficients of the Holt–Winters residuals are slightly lower than those of the adaptive regression at most lags. Although none of the methods completely eliminates the weekly correlation effect, but the Holt–Winters method comes closest. Note the low Holt–Winters 1-day lag coefficient compared to those of the regression residuals.

Table II gives the values of τ_1 and τ_7, the autocorrelation coefficients for all 16 series at lags of 1 day and 7 days, respectively. The Holt–Winters method shows consistently low autocorrelation

DOI: 10.1002/sim

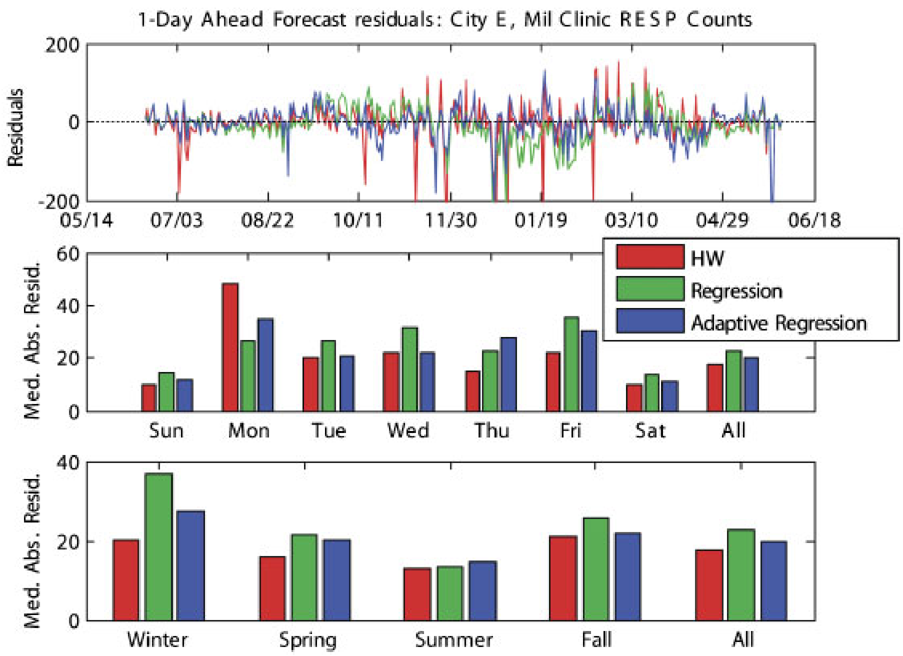

Plate 3. Comparison of forecast residuals startified by day-of-week and by season.

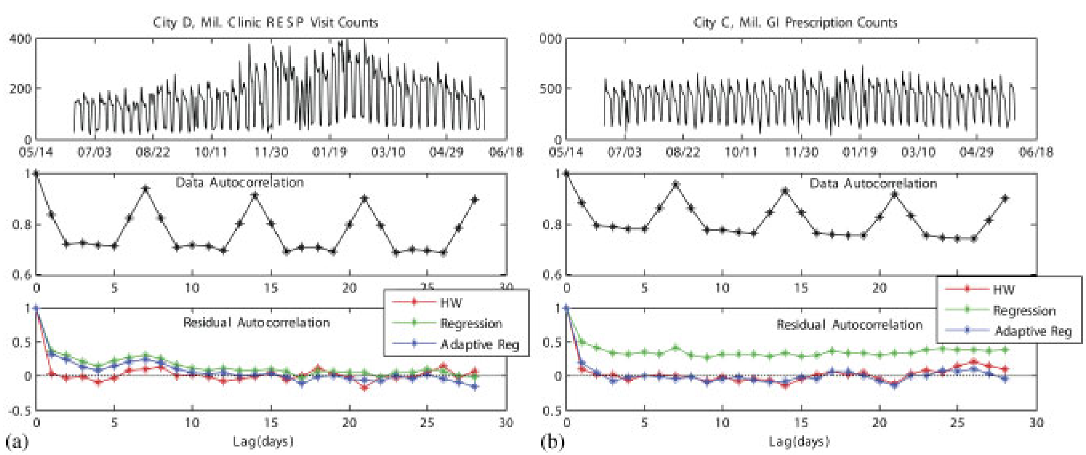

Plate 4. Data and residual autocorrelation coefficients for time series of: (a) respiratory and (b) gastrointestinal syndrome counts.

DOI: 10.002/sim

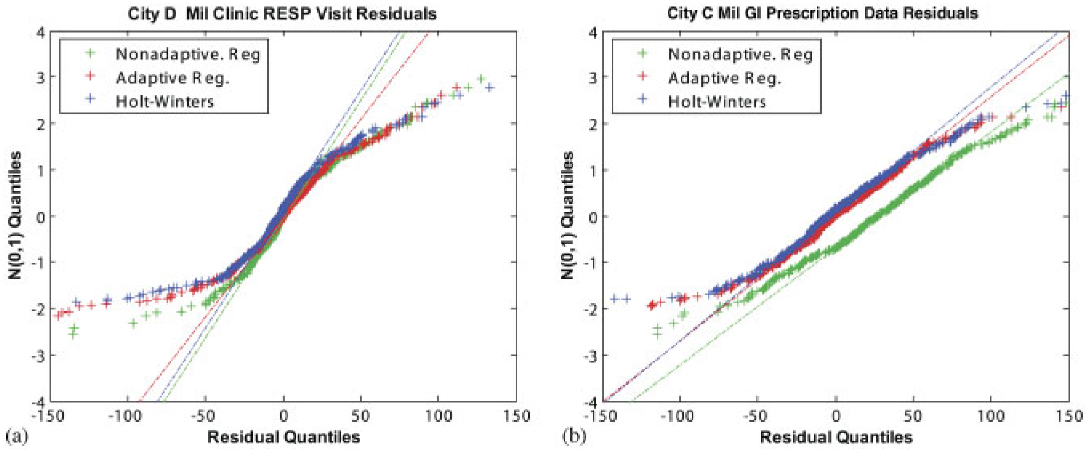

Plate 5. Residual distributions shown as normal probability plots for: (a) RESP data for military clinic visit counts for City D and (b) GI data for military GI prescription sales for City C.

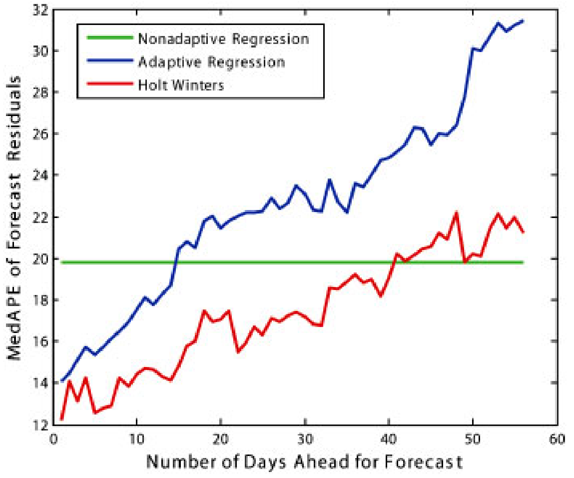

Plate 6. Degradation of adaptive forecast methods *versus* non-adaptive regression as a function of prediction days ahead.

Table II. Comparison of autocorrelation coefficients for 1-day-ahead forecast residuals at lags of 1 and 7 days.

	$\tau_1 = 1$-day lag coefficient			$\tau_7 = 7$-day lag coefficient		
Time series	Non-adaptive regression	Adaptive regression	Holt–Winters	Non-adaptive regression	Adaptive regression	Holt–Winters
Autocorrelation coefficients for 1-day-ahead forecast residuals						
RESP 1	0.59	0.36	**0.11**	0.62	0.26	**0.12**
RESP 2	0.58	0.28	**0.14**	0.60	0.17	**0.10**
RESP 3	0.76	0.21	**0.10**	0.79	**0.03**	0.06
RESP 4	0.37	0.32	**0.03**	0.31	0.24	**0.09**
RESP 5	0.50	0.37	**0.13**	0.40	0.23	**0.09**
RESP 6	0.67	0.34	**0.24**	0.66	**0.06**	0.13
RESP 7	0.53	0.28	**0.08**	0.57	0.13	**0.07**
RESP 8	0.58	0.34	**0.13**	0.56	**0.19**	0.20
RESP 9	0.76	0.59	**0.22**	0.65	0.43	**0.16**
RESP 10	0.79	0.38	**0.16**	0.91	0.26	**0.17**
GI 1	0.49	0.13	**0.13**	0.55	0.05	**−0.02**
GI 2	0.67	0.09	**0.05**	0.67	0.05	**0.00**
GI 3	0.75	0.24	**0.14**	0.75	0.07	**0.06**
GI 4	0.50	0.20	**0.11**	0.42	−0.05	**0.01**
GI 5	0.63	0.22	**0.18**	0.65	**0.02**	0.09
GI 6	0.58	**0.16**	0.24	0.62	0.04	**0.02**

Note: Bold values signify the lowest coefficient among the three methods for the corresponding data series.

at these lag values. In [15], Chatfield reported higher 1-day-lag coefficients τ_1 for Holt–Winters residuals from time series with different features and gave a method for further reducing the autocorrelation by using τ_1 as another smoothing coefficient. He recommended this stratagem if τ_1 exceeds $2/\sqrt{N}$, where N is the residual series length. Most of the Holt–Winters τ_1 values computed here were only slightly above this threshold, so we did not attempt this improvement.

Plate 5 displays normal probability plots for two of the cities. It appears that the majority of the residuals, for all three methods, fall close to the straight line thereby indicating normality. The fact that the negative tails seem to diverge further from the line than the positive tails is most likely and unsurprisingly the result of overprediction on holidays. Similar patterns were obtained for the remaining series. The determination of alerting algorithm thresholds should take account of these long distribution tails.

4.3. Comparison of residuals for 7-day ahead predictions

We also compared the extended forecast capabilities of these methods because derived anomaly detection algorithms may be used to seek the data signature of both sudden and gradual outbreaks. A general biosurveillance capability requires sensitivity to both types of signal, depending on the distribution of the disease incubation period and on the data acquisition rate. The use of 1-day-ahead forecasts could mask a gradual data signal because its early cases could increase the forecasts and cause the subsequent observed counts to appear within expected limits. We therefore repeated the method comparisons for 7-day-ahead forecasts.

DOI: 10.1002/sim

Table III. Comparison of 7-day-ahead forecasts for 10 respiratory count series and six gastrointestinal count series, using median absolute residual, median absolute per cent error, and root-mean-square residual.

	Respiratory count series	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
MAD	Non-adaptive regression	22.6	31.7	83.1	**13.3**	**27.3**	35.6	43.9	53.7	76.7	131.1
	Adaptive regression	**20.4**	36.9	28.1	18.0	35.8	23.5	32.0	60.0	81.8	49.3
	Holt–Winters	21.1	**28.9**	**26.0**	15.3	28.6	**22.5**	**24.7**	**39.2**	**63.7**	**35.3**
MedAPE	Non-adaptive regression	11.7	11.6	18.8	**10.0**	**10.5**	19.1	19.8	8.6	10.4	32.6
	Adaptive regression	11.1	11.8	7.3	15.0	14.1	15.5	16.1	9.6	12.2	13.0
	Holt–Winters	**9.0**	**9.3**	**6.3**	12.3	11.8	**13.2**	**12.9**	**6.6**	**9.6**	**10.1**
RMSE	Non-adaptive regression	44.8	69.3	110.0	**34.3**	**71.2**	74.6	65.1	113.3	**197.9**	244.1
	Adaptive regression	**44.7**	**67.9**	**76.6**	45.0	98.7	62.8	**61.1**	122.1	259.7	141.6
	Holt–Winters	55.9	89.9	79.3	50.9	104.8	**59.1**	64.8	**109.1**	218.0	**104.3**
	GI count series	(a)	(b)	(c)	(d)	(e)	(f)				
MAD	Non-adaptive regression	13.5	35.6	24.3	38.9	24.1	21.1				
	Adaptive regression	9.4	16.9	10.5	25.0	13.4	13.3				
	Holt–Winters	**8.9**	**16.7**	**8.9**	**23.4**	**10.7**	**12.1**				
MedAPE	Non-adaptive regression	15.3	17.8	23.4	10.7	22.6	20.5				
	Adaptive regression	11.3	**8.2**	10.3	7.4	12.0	13.7				
	Holt–Winters	**10.1**	8.2	**8.4**	**6.8**	**8.8**	**11.7**				
RMSE	Non-adaptive regression	19.7	51.5	32.2	**58.3**	43.5	32.9				
	Adaptive regression	**17.2**	**41.4**	22.5	61.4	34.7	**27.0**				
	Holt–Winters	19.2	55.8	**22.0**	78.2	**33.6**	28.7				

Note: Bold values signify lowest residual measure for that city among the three methods.

Making extended forecasts was straightforward with the methods of this study. For the non-adaptive regression, no change was necessary because the model coefficient estimates depend only on the training period, so no recalculation was necessary. For the adaptive regression, we inserted a sliding 7-day buffer period between the 8-week baseline and the forecast day so that a model fit could not use knowledge of data within a week of the forecast. For the Holt–Winters method, we simply set $k=7$ in forecast equation (5).

Table III presents the MAD, MedAPE, and RMSE comparisons of the 7-day-ahead residuals, analogous to the 1-day-ahead results in Table I. The non-adaptive regression residuals are unchanged. For the RESP series with more trend-like effects because of the seasonality, the adaptive entries show approximately 10–20% degradation from Table I, depending on the measure. For the GI series, the degradation is slight, especially for the Holt–Winters method. However, both adaptive methods still significantly outperform the non-adaptive regression overall, and for none of the series did the 7-day forecast extension cause an advantage for the non-adaptive method by any measure.

Plate 6 shows the resulting comparison plot for forecast lags of up to 56 days. The performance advantage of non-adaptive over adaptive regression persists for two weeks, which is longer than the expected spread of incubation periods of most diseases known to be weaponized for bioterrorism [25]. The explicit trend term in the adaptive regression causes increasing errors for longer forecast lags. The advantage of the Holt–Winters predictions over the non-adaptive ones persists up to six weeks of forecast lag.

DOI: 10.1002/sim

5. DISCUSSION

We performed this study to compare forecast methods for biosurveillance data. Robust forecast methods are required to obtain residuals for use as input to alerting algorithms. For reliable sensitivity at manageable false alarm rates, such algorithms require input data as free as possible of trends, cyclic effects, and other sources of autocorrelation. Given the evolving, non-stationary nature of these data, based on the care-seeking behaviour of a changing population, obtaining such series is a challenging problem.

To obtain the required forecasts, a number of authors have applied regression models commonly used in the retrospective studies of biostatistics. For reliable prospective utility, these models require some type of modification because the statistical behaviour in one time interval is used to predict the behaviour in a later interval, and this behaviour commonly changes. Longer model-training periods do not guarantee better forecasts, and they are often unavailable anyway in these data streams.

Researchers have suggested model adjustment methods of varying complexity, such as refitting the model on a regular basis or when the some goodness-of-fit criterion fails. Autoregressive strategies have been successfully applied to build adaptive capability into the modelling. Given a syndromic data source in a 'steady state' with adequate history and the on-hand expertise to select an appropriate ARIMA model, adaptive forecasts may be obtained. However, selection of appropriate ARIMA models requires analysis of each data type. This process may not be practical for an automated biosurveillance system that includes many evolving time series and needs to periodically add new data sources and rapidly model new syndrome categories. We restricted our forecast methods to automatic implementations not requiring analysis of data stream characteristics.

We applied the Holt–Winters method for generalized exponential smoothing as an alternative to data modelling. This method has several advantages. It is easy to understand and apply for a large class of data types. Only a few modifications were required for practical implementation. Little data history is needed. Since this method is highly adaptive and well suited to capturing local trends, it avoids the need for artificial adaptations that may have to change across data sources and syndrome groupings. A limitation is that generalized exponential forecasting cannot directly exploit external information such as the daily temperature or the number of participating data providers without modification. Regression models can include this information with additional covariates.

For our forecast comparisons, we chose 16 city-level syndromic time series with distinct day-of-week effects because these series seemed well suited to a recent loglinear regression approach published by Brillman *et al*. [11]. Ten of these series also displayed seasonal trends. The three forecast methods we compared were a regression model using the first 350 days of each series as a fixed baseline, an adaptive regression model with a sliding 8-week baseline, and Holt–Winters exponential smoothing. When we discounted the effects of 10 calendar holidays, the non-adaptive regression yielded the largest forecast residuals, and in a few instances these were obvious over entire seasons. For both 1-day ahead and 7-day ahead prediction, the Holt–Winters forecasts were slightly but consistently closer to the observed counts than those of the adaptive regression.

Regarding the forecast methods as preconditioning filters for control charts, the effects of the two adaptive methods were encouraging. The 1-day and 7-day-lag autocorrelation coefficients τ_1 and τ_7 were over 0.85 and 0.95 for all 16 data series, respectively, and they were sharply reduced in the residual series. For the Holt–Winters residuals, the τ_1 values were all below 0.25, and they were below 0.15 for 11 of the 16 series. None of the τ_7 values exceeded 0.2, and for the six non-seasonal series, all were below 0.1. Thus, this automatic smoothing implementation produced series approaching the temporal independence assumed for control charts. The use of such series

DOI: 10.1002/sim

can allow examination of alerting algorithms without confusion by the autocorrelation of the inputs. These residuals can also enable research in multivariate alerting methods, though for that work the correlation across input data streams must also be understood.

This study suggests several research issues:

- An important problem for the biosurveillance data environment is the classification of data streams for appropriate model selection. For regression modelling, this classification is needed to choose covariates and to decide when the model must be changed. For generalized exponential smoothing, the classification is needed for appropriate selection of smoothing coefficients. Our study suggests that this smoothing is more robust to its smoothing coefficients than are regression models to their covariates. This observation and the study results in general need further testing using time series of the different scales and seasonal behaviours that can result from geographic, syndromic, and temporal aggregation decisions. A comprehensive data classification strategy will require this additional testing.
- Another issue is the management of expected and unexpected data outliers. Management questions include both alerting and training based on these outliers. Regression modelling can treat expected outliers using covariates such as holiday indicator variables. For unexpected outliers, we have implemented automated outlier removal schemes to avoid baseline contamination for the adaptive regression, but such schemes can produce unexpected effects and need further study. For the Holt–Winters method, Section 2.3 describes our changepoint detection criterion for skipping the update equations to avoid improper training, but we made no modifications to avoid alerting on known holidays. Further adaptations could improve the detection performance of this method.
- For some data sources, additional information such as the daily number of total clinic visits of any type or of all sales may be available, and external information such as daily temperature or air quality measurements may be useful. The question becomes how to use this knowledge to improve forecasts. For the regression modelling, the information may be included as additional covariates. For Holt–Winters and other smoothers based only on internal counts, this information can be integrated in several ways: one is to create a new time series that combines the daily counts with external information. For example, exponential smoothing could be applied to the ratio of respiratory visits to all visits. More complex or multivariate information could be included in a two-stage process, removing the effects of external covariates with data modelling and then applying the smoothing to those residuals. Finally, Pfeffermann and Allon introduced a method for improving forecasts of one series by integrating information from other time series using a multivariate Holt–Winters formulation [26]. In their study, they produced forecasts more accurate than those from the univariate series alone and even compared to an ARIMA model. However, the method requires the specification of a matrix of smoothing parameters and their initial conditions, which adds a non-trivial level of complexity.
- While robustness across data types was not a primary focus of this methodological study, the observations about the relative performance of the forecast methods were consistent across both syndrome groups and data sources, to the extent that clean data were available. For example, the respiratory syndrome study data contained multiple time series for military clinic visits, prescription sales, and civilian physician office visits, and the forecast conclusions were the same for each data type. No contradictions were found among the smaller set of GI-based series. More data sets of varying geographic scale are needed for stronger evidence of robustness.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

In preparation of this document, the first and second authors were partially funded under grant #8AUAS5 from the Office of State and Local Government Coordination and Preparedness (SLGCP), United States Department of Homeland Security. Points of view or opinions expressed in this document are those of the authors and do not necessarily represent the official position or policies of SLGCP or the U.S. Department of Homeland Security.

REFERENCES

1. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. *Emerging Infectious Disease* 1997; **3**:395–400.
2. Williamson GD, VanBrackle G. A study of the average run length characteristics of the National Notifiable Diseases Surveillance System. *Statistics in Medicine* 1999; **18**(23):3309–3319.
3. Ryan TP. *Statistical Methods for Quality Improvement*. Wiley: New York, 1989.
4. Benneyan JC. Part I: introduction and basic theory, part II: chart use, statistical properties, and research issues. *Statistical Quality Control Methods in Infection Control and Hospital Epidemiology*, *Infection and Hospital Epidemiology* 1998; **19**(3):194–214.
5. Shmueli G, Fienberg SE. Current and potential statistical methods for monitoring multiple data streams for bio-surveillance. In *Statistical Methods in Counter-Terrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication,* Wilson A, Wilson G, Olwell DH (eds). Springer: Berlin, 2006. ISBN: 0-387-32904-8.
6. Burkom HS, Elbert Y, Feldman A, Lin J. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. *Syndromic Surveillance*: *Reports from a National Conference*, 2003. *Morbidity Mortality Weekly Report* 2004; **53**(Suppl.):67–73.
7. Hurt-Mullen KJ, Coberly J. Syndromic surveillance on the epidemiologist's desktop: making sense of much data. *MMWR Morbidity Mortality Weekly Report 2005* 2005; **54**(Suppl.):141–146.
8. Woodall WH. The use of control charts in health care monitoring and public health surveillance. *Journal of Quality Technology* 2006; **38**(2):89–105.
9. Mandel BH. The regression control chart. *Journal of Quality Technology* 1969; **1**(1):1–9.
10. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology* 2004; **159**:217–224.
11. Brillman JC, Burr T, Forslund D, Joyce E, Picard R, Umland E. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance. *BMC Medical Informatics and Decision Making* 2005; **5**(4):1–14. http://www.biomedcentral.com/content/pdf/1472-6947-5-4.pdf
12. Burkom HS. Development, adaptation, and assessment of alerting algorithms for biosurveillance. *Johns Hopkins APL Technical Digest* **24**(4):335–342.
13. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making* 2003; **3**:2.
14. Chatfield C. The Holt–Winters forecasting procedure. *Applied Statistics* 1978; **27**:264–279.
15. Chatfield C, Yar M. Holt–Winters forecasting: some practical issues. *The Statistician* 1988; **37**:129–140.
16. Lombardo JS, Burkom HS, Elbert YA, Magruder SF, Lewis SH *et al*. A systems overview of the electronic surveillance system for the early notification of community-based epidemics. *Journal of Urban Health*, *Proceedings of the 2002 National Syndromic Surveillance Conference*, New York, NY, vol. 80, No. 2, Supplement 1, April 2003; i32–i42. http://jurban.oupjournals.org
17. Brown RG, Meyer RF. The fundamental theorem of exponential smoothing. *Operations Research* 1961; **9**:673–685.
18. Stoto MA, Schonlau M, Mariano LT. Syndromic surveillance: is it worth the effort? *Chance* 2004; **17**:19–24.
19. Hogan WR, Tsui FC *et al*. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. *Journal of the American Medical Information Association* 2003; **10**(6):555–562.
20. Albright SC, Winston WL, Zappe CJ. *Data Analysis for Managers* (2nd edn). Duxbury Publishing: Belmont, CA, U.S.A., 2004.
21. Holt CC. Forecasting seasonals and trends by exponentially weighted averages. *ONR Memorandum 52/1957*, Carnegie Institute of Technology.
22. Winters PR. Forecasting sales by exponentially weighted moving averages. *Management Science* 1960; **6**:324–342.

23. *NIST/SEMATECH e-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc435.htm, accessed 10 November 2006.
24. Siegrist D, Pavlin J. BioALIRT biosurveillance detection algorithm evaluation. *Syndromic Surveillance*: *Reports from a National Conference*, 2003. *MMWR* 2004; **53**(Suppl.):152–158.
25. U.S. Army Medical Research Institute of Infectious Diseases. *Medical Management of Biological Casualties*. U.S. Army Medical Research Institute of Infectious Diseases: Ft. Detrick MD, September 2000.
26. Pfeffermann D, Allon J. Multivariate exponential smoothing: method and practice. *International Journal of Forecasting* 1989; **5**(1):83–98.

DOI: 10.1002/sim