Markov Switching Models for Disease Outbreak Detection

1 INTRODUCTION

Detecting and controlling infectious disease outbreaks have long been a major concern in public health [1]. Recent efforts in building syndromic surveillance systems have included increasing the timeliness of the data collection process by incorporating novel data sources such as emergency department (ED) chief complaints (CCs) and over-the-counter (OTC) health product sales [2]. Research shows that these data sources contain valuable information that reflects current public health status [3], [4], [5], [6]. However, the noise caused by routine behavior patterns, seasonality, special events, and various other factors is blended with the disease outbreak signals. As a result, disease outbreak detection using the time series from syndromic surveillance systems is a challenging task.

In a typical syndromic surveillance system [7], [8], [9], [10], [11], [12], the data are classified and aggregated to generate univariate or multivariate time series at a daily frequency. An example of a univariate time series is the daily ED visits associated with a particular syndrome (for example, the respiratory syndrome). An example of a multivariate time series is the number of daily visits with a particular syndrome from multiple EDs. If geographic information such as the ZIP code is available, the multivariate time series in these examples would be the daily counts of patients with a particular syndrome from the ZIP code areas near an ED.

Most time series outbreak detection methods follow a two-step procedure [13], [14], [15], [16]. In the first step, a baseline model describing the "normal pattern" is estimated using the training data that usually contain a historical time series without outbreaks. The baseline model then is used to predict future time series values. In the second step, statistical surveillance methods such as the Shewhart control chart [17], [18] or the Cumulated SUM (CUSUM) [19] method then take the prediction error (observed value minus predicted value) as the input, and output alert scores. Higher alert scores are usually associated with a higher risk of having outbreaks. When the alert scores exceed a predefined threshold, the alarm is triggered.

Two main problems exist for current detection methods. First, the two-step procedure is based on the assumption that there are no outbreaks in the training data. When a real-world dataset is used for training, the assumption is very hard to verify. Moreover, a full investigation of disease outbreaks during the data collection period is usually too expensive to conduct.

The validity of the detection results may be seriously impaired if it cannot be verified that the training data are outbreak-free. The estimated parameters of the baseline model may be biased by outbreak-related observations. Subsequent prediction and outbreak detection, as a result, may be negatively affected. The problem can seriously reduce the practical value of the outbreak detection method.

Second, existing time series detection methods also lack the ability to handle sporadic extreme values. Special events such as holidays and the media coverage of a particular disease may cause spikes that are not associated with disease outbreaks [20]. These extreme values usually last for a very short time (often just one or two days) and do not affect subsequent time series values. Anomalies related to real disease outbreaks, on the other hand, usually show a prolonged upward drift. The magnitude of disease-related drift is usually much smaller compared to the sporadic spikes caused by special events. Many outbreak detection algorithms take advantage of these characteristics and accumulate the errors so that small increases can be detected effectively [14], [15], [21]. The accumulation process, nevertheless, is susceptible to the presence of extreme values.

The deficiencies of current outbreak detection methods motivate our efforts to develop novel algorithms that can address these shortcomings. To deal with the problem of having outbreak-related observations in training data, a flexible statistical model must be used so that the model can adjust itself automatically when outbreak-related observations exist. In econometrics and time series literature, this is usually refereed to as the problem of modeling endogenous structural changes [22], [23].

A natural way of modeling structure changes in a time series is introducing additional hidden state variables which control the underlying time series dynamics. The Markov switching models originally proposed by Hamilton are one popular model of this kind [24]. This family of models includes a hidden state variable that may have a different value in each period. It takes values of either 0 or 1 that correspond to different conditional means, variances, and autocorrelations of the time series. The hidden state evolves following a first-order Markov process. That is, the current hidden state depends only on its historical values from the last period.

This hidden state method can be easily extended to handle extreme values. An additional hidden state can be included to model the presence of sporadic extreme values. With this additional hidden state, the model can distinguish between “normal” and “extreme” observations. That is, if a spike appears without signs that the sudden increase can be associated with drifts either before or after it, then the model can, based on the statistical evidence, assign the sudden increase as an extreme value instead of an outbreak. The negative effect of extreme values on outbreak detection can thus be reduced.

The main contribution of this paper is to present a prospective outbreak detection method that is robust to pre-existing outbreaks and extreme values. Prospective outbreak detection, as opposed to retrospective detection in which the entire set of the observations is available to the detection algorithms, assumes that only observations made prior to the time of the detection are available to the detection algorithms. Retrospective detection is useful primarily for offline analysis of historical data, whereas prospective detection is intended for use in monitoring incoming public health data streams in an online fashion. We utilized the Markov switching model that includes three hidden state variables in each period. The first hidden state variable models the disease outbreak state and the second hidden state variable models the presence of extreme value. If the extreme value exists, the third hidden state variable represents the size of the extreme value. We demonstrate that our approach

outperforms several existing state-of-art outbreak detection algorithms using both simulated and real-world time series data.

This paper is organized as follows. Section 2 briefly introduces current outbreak detection methods and the Markov switching models. Section 3 presents our outbreak detection method. An evaluation study that uses both simulated and real-world data is summarized in Section 4. We conclude our paper in Section 5.

2 BACKGROUND

Current time series outbreak detection methods mostly follow a two-step procedure: a base-line time series estimation step followed by a statistical surveillance step [13], [14], [15]. We review these two major steps in this section.

Markov switching models, which belong to a broader class of statistical models that make use of hidden state variables, are also reviewed. We present the typical model settings and the estimation approaches.

2.1 Time Series Modeling

The first step in traditional outbreak detection methods is to develop a model that can describe the normal time series patterns. The most widely used model is the Autoregressive Integrated Moving Average (ARIMA) models of Box and Jenkins [25]. The model setting can be described by three parameters: (p,d,q). The parameter p refers to the length of historical time series values that can affect current observations. The second parameter d specifies how many difference operations are required to make the time series stationary. The third parameter q specifies the length of historical error terms that can affect current observations. In a typical setting that does not involve seasonal fluctuation, the observed time series is usually assumed to be stationary, that is, $d = 0$. Specifically, an ARIMA(p,0,q) model can be written as:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} \cdots + b_q \varepsilon_{t-q}$$

where y_t is the observed time series and ϵ_t is the error term. To ensure that the model "learns" the normal time series pattern, the data used for model estimation should be outbreak free. Given p and q, the parameter values (a_0, a_1, ..., b_q) can be estimated using likelihood maximization [26]. However, different model settings that correspond to different values of p and q may affect prediction accuracy. The values of p and q are usually determined by model selection criteria that take both goodness of fit and model complexity into consideration. Commonly used model selection criteria include Akaike information criterion (AIC) [27], [28] and Bayesian information criterion (BIC) [29]. Note that the model selection criteria are closely related to the "cross-validation" evaluation approach [30] commonly used by the machine learning community [31]. In fact, cross-validation is asymptotically equivalent to AIC [32].

Other modeling techniques such as the generalized linear model using Poisson distribution [33], expectation-variance model [34], and the Wavelet Model [35] have been evaluated in previous studies.

For the purpose of detecting outbreaks, there are two issues warranting further discussion: the modeling of the day-of-week and seasonal effects.

2.1.1 Day-of-Week Effect

The syndromic surveillance time series usually exhibits strong day-of-week effects. For example, there are usually more ED visits during the weekends than during the weekdays [15]. The variation among different day-of-weeks is usually assumed to be fixed. As an illustrative example, an ARIMA(1,0,0) model with a fixed day-of-week effect can be written as

$$y_t = w_1 d_{t,1} + w_2 d_{t,2} + \cdots + w_6 d_{t,6} + a_0 + a_1 y_{t-1} + \varepsilon_t$$

where $d_{t,i} \in \{0,1\}, i = 1,2,3,4,5,6$ are dummy variables indicating a particular day-of-week. For example $d_{t,1} = 1$ if day t is a Monday and 0 otherwise. Note that we need only 6 dummy variables for 7 day-of-weeks because of the existence of the constant term a_0.

2.1.2 Seasonal Effect

Similar to the day-of-week effect that refers to a weekly cyclic pattern, the seasonal effect refers to a yearly cyclic pattern. Tri-geometric functions are commonly used to model deterministic seasonal fluctuation. This technique is usually referred to as the Serfling model [36], [37], which can be written as:

$$y_t = a_0 + b_1 \cos\left(\frac{2\pi t}{365.25}\right) + b_2 \sin\left(\frac{2\pi t}{365.25}\right) + w_1 d_{t,1} + w_2 d_{t,2} + \cdots + w_6 d_{t,6} + \varepsilon_t$$

Note that both day-of-week and seasonality are included in the model. The model can be refined by including more tri-geometric functions that correspond to semi-annual and even quarterly cyclic patterns. However, it has the obvious problem of assuming the same seasonal peaks and troughs across the whole monitoring period [37]. Our preliminary experiments show that the Serfling model fits the observed syndromic time series poorly especially when the seasonality is strong. The Serfling model assumes a particular shape of the time series that may not be empirically valid.

Other modeling techniques allow more flexible seasonal fluctuation across years. One possibility is to use the Holt-Winters exponential smoothing to model seasonality [38], [39]. An empirical study showed that, in the context of syndromic surveillance, Holt-Winter exponential smoothing outperformed the Serfling model in terms of prediction accuracy [40].

The concept of the seasonal random walk [41] can be applied to model the seasonal effect. The basic idea is that the same day-of-year should have the same expected value. Reis and his colleagues estimated the expected value using the trimmed-mean of historical time series value with the same day-of-year in an 8-year window [14], [15]. The seasonal effect can then be filtered out by subtracting the observed value from the day-of-year expectation.

2.2 Statistical Surveillance Methods

For outbreak detection purposes, the prediction errors from the time series modeling step are further processed using statistical surveillance methods. Various statistical surveillance methods such as the Shewhart control Chart [17], Cumulated Sum (CUSUM) [19], Exponential Weighted Moving Average (EWMA) [18], Shiryaew-Roberts method [42], [43] and the likelihood ratio methods [44] can be applied for disease outbreak detection. However, most syndromic surveillance studies use the Shewhart control chart, CUSUM, EWMA and their variations. Our review focused mainly on these three methods. More detailed reviews can be found elsewhere [45].

The Shewhart control chart [17] checks the t-value of the prediction errors period by period. It performs the best if large, isolated outbreaks are involved. However, since disease outbreaks often exhibit only small deviations in their early stages, the Shewhart control chart may not be the best choice for our purposes.

The CUSUM method minimizes the maximum value of the conditional expected delay "when the outcome before outbreak is the worst possible" [46]. It uses a recursive formula to accumulate the prediction errors:

$$C_t^+ = \max\left[0, e_t - K + C_{t-1}^+\right]$$

where e_t is the prediction error from the time series model and K is a predefined constant that is commonly referred to as the allowance. The alarm is triggered if C_t^+ exceeds a predefined threshold.

The EWMA method can be seen as a linear approximation of the likelihood ratio method [44], [47]. The alert score is computed by accumulating forecasting errors with exponentially decaying weights. Similar to the CUSUM method, higher outbreak scores are usually associated with a higher risk of having an outbreak. The threshold can be determined from theoretical analysis or empirical studies [48], [49].

Some syndromic surveillance studies use a moving average scheme to accumulate forecasting errors [14], [15]. Their studies have showed that a linear increasing weighting schemes performed best in terms of outbreak detection ability.

2.3 Performance Measures

The most commonly used performance measure in statistical surveillance literature is the Average Run Length (ARL). ARL^0 denotes the expected run length until the first false alarm, and ARL^1 denotes the expected run length until an alarm when the process is out of control at the start of the surveillance [45], [50], [48], [49].

These measures, nevertheless, are less intuitive under the context of disease outbreak detection. Most disease outbreak detection studies use per day sensitivity and false alarm rate [33], [34], [14]. Sensitivity is the probability of having alarms on outbreak days. False alarm rate is the probability of having alarms on non-outbreak days.

2.4 Extreme Values in Syndromic Surveillance Time Series

Current surveillance methods are very sensitive to extreme values. The main reason is because the statistical surveillance methods accumulate the forecasting errors and there are no simple methods that can be used to filter out the extreme values. Burkom [51] proposed using a "reset" rule to bring down the alert scores when extreme values are known to be causing the elevated scores. However, it is not clear how to establish effective reset rules.

Common reasons behind the extreme values include holidays, media coverage, and special events [20]. However, existing studies have not offered help for handling the negative effects caused by the extreme values. Previous studies have used holiday dummies to absorb the holiday effects [34]. This technique, nevertheless, imposes an unrealistic assumption that all holidays have the same effect on the time series.

2.5 The Markov Switching Model

The Markov switching model belongs to the family of state-space models. A state-space model is a statistical model with hidden state variables controlling observable random variables. There are two types of equations in this model: the measurement equations and the transition equations [52]. The measurement equation defines how hidden states affect the observable random variables. The transition equation, on the other hand, defines how the state variables evolve over time.

When the state variable is discrete, the state-space model is usually called the hidden Markov model [53], [54] or the Markov switching model [24] depending on the choice of the measurement equation. The measurement equation in the hidden Markov model is usually formulated so that the observable random variables at period t only depend on the hidden state variables at the same period.

The Markov switching model addresses the weakness of the hidden Markov model by including lagged observations. The observable random variables in the Markov switching model depend on their historical values as well as the hidden state variables. This setting makes the Markov switching model more suitable for time series related problems. Figure 1 illustrates the dependency difference between the Markov switching model and the hidden Markov model.

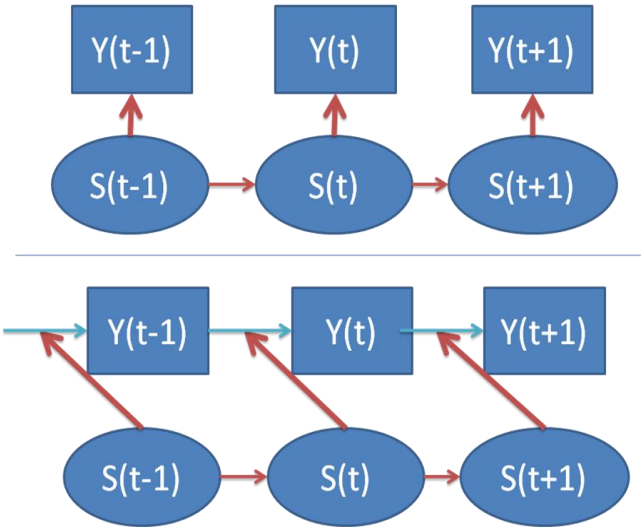

Fig. 1. Markov Switching models (upper panel) and hidden Markov models (lower panel). The rectangles are observable random variables and the circles are hidden state variables. Arrows indicate the dependencies among variables.

Strat and Carrat [55] applied the state-space model for disease outbreak detection. They used a two-state hidden Markov model on a weekly influenza-like illness (ILI) incidence and showed that the hidden Markov model clearly differentiated between epidemic and non-epidemic rates. However, as they pointed out in the conclusion, "the validity of the hypothesis that ILI incidence rates are independent conditional on the state is questionable." They also pointed out that autoregressive terms should be included for better performance. We are unaware of prior studies on applying Markov switching models for outbreak detection.

Most applications of the Markov switching models fall in the field of economics and finance. Notable examples are identifying macroeconomics business cycle [24] and modeling changing interest rates regimes [56]. A simple Markov switching model can be written as

$$y_t = a_{0,0} + a_{0,1}s_t + (a_{1,0} + a_{1,1}s_t)y_{t-1} + e_t \quad (1)$$

$$p(s_t = j \mid s_{t-1} = i) = p_{ij} \quad (2)$$

$$s_t \in \{0,1\} \quad (3)$$

$$e_t \sim N(0, \sigma^2) \quad (4)$$

Equation 1 defines how the hidden state variable s_t controls the dynamics of the observable random variable y_t. At an non-outbreak period ($s_t = 0$), y_t is determined by a drift term $a_{0,0}$ and the autoregressive parameter $a_{1,0}$. If an outbreak occurs ($s_t = 1$), the drift term increases to $a_{0,0} + a_{0,1}$ and the autoregressive parameter increases to $a_{1,0} + a_{1,1}$ (assuming $a_{0,1} \geq 0$ and $a_{1,1} \geq 0$). Equation 2 indicates that the hidden states evolve following a Markov process with transition probability p_{ij}.

Note that if we have a time series of T period, there are 4 parameters and T hidden state variables in Equation 1, together with 2 variables for transition probability in Equation 2 and a variance for error terms in Equation 4. We have more unknowns than the number of periods, which complicates the estimation process. We briefly discuss the model estimation issues below.

2.5.1 Model Estimation for the Markov Switching Model

Model estimation for the Markov switching model is much more complicated than that of the standard time series models such as the ARIMA models. The technical difficulty arises from the presence of unknown hidden states. In a simplified case involving only one hidden outbreak state variable with two possible states and a total of T periods, a direct evaluation of the likelihood function involves a summation of all possible trajectories of hidden states. The time complexity is $O(2^T)$, which is intractable in practice. More sophisticated algorithms, which compute the posterior distribution of the hidden states using a forward-filtering-backward-smoothing (FFBS) procedure [52], [57], take only $O(2^3T)$ steps. The computation of the posterior distribution of the hidden states is required by many estimation methods such as the expectation-maximization (EM) algorithm [58], [59], [60], Gibbs sampling, and Markov Chain Monte Carlo (MCMC) [61], [62], [63]. Note that to deliver the final optimal parameter estimation, these algorithms need to execute repeatedly until certain convergence criteria are met.

The EM algorithm finds the maximum of the likelihood function by iterating between calculating the expected value of state variables given current parameters and calculating the maximum of log likelihood given the expected state variables. It was applied to estimate the hidden Markov model in a previous outbreak detection study [55]. Compared to other numerical optimization methods, the EM algorithm is more robust and usually converges if a maximum exists. However, it is possible that the algorithm converges to a local maximum instead. In practice, the EM algorithm is run with multiple initial values.

A serious drawback of the EM algorithm is the label switching problem [57]. The Markov switching model (and the hidden Markov model) is invariant under arbitrary permutations of the state labels. As a result, we cannot be sure whether $s_t = 0$ is representing an outbreak or non-outbreak state before the estimation procedure is completed. The label switching problem is especially an issue when the Markov switching model is part of a larger automatic disease outbreak detection system.

Gibbs sampling [62], [63], [64] is an alternative estimation method that can avoid the label switching problem. The Gibbs sampling iterates to draw random variables from conditional posterior distributions of parameters and state variables to simulate the full posterior distribution of parameters and state variables. Specifically, let $\Theta = \{\theta_1,...,\theta_k\}$ denote the unknown parameters (and state variables). By the Bayes Theorem, the posterior distribution $p(\Theta|Y)$ is proportional to the likelihood of $p(Y|\Theta)$ multiplying the prior of parameters $p(\Theta)$. The label switching problem can be avoided by imposing proper constraints on $p(\Theta)$. Gibbs sampling estimates parameters using a simulation-based method. The following steps can be used to simulate Θ from its posterior distribution. First, select initial values $\Theta^{(0)} = \{\theta_1^{(0)},...,\theta_k^{(0)}\}$. For $i = 1,2,...I$, iterate through the following steps:

1. Draw $\theta_1^{(i)}$ from $p(\theta_1|Y,\theta_2^{(i-1)},...,\theta_k^{(i-1)})$.
2. Draw $\theta_2^{(i)}$ from $p(\theta_2|Y,\theta_1^{(i)},\theta_3^{(i-1)},...,\theta_k^{(i-1)})$.
   ...
3. Draw $\theta_k(i)$ from $p(\theta_k|Y,\theta_1^{(i)},\theta_2^{(i)},...,\theta_{k-1}^{(i)})$.
4. Record $\Theta^{(i)} \equiv \{\theta_1^{(i)},\theta_2^{(i)},\ldots,\theta_k^{(i)}\}$

It has been shown that $\{\Theta^{(i)}\}$ converges to $p(\Theta|Y)$ [65], [66]. As a result, the posterior mean of θ_j can be estimated by the average of $\{\theta_j^{(i)}\}$, excluding certain "burn-in" iterations to minimize the effect of the initial value. The confidence intervals of the estimated parameters can also be calculated directly from $\{\theta_j^{(i)}\}$.

3 OUTBREAK DETECTION USING THE MARKOV SWITCHING WITH JUMPS (MSJ) MODEL

We developed our disease outbreak detection algorithm based on the Markov switching models [24]. Two hidden disease outbreak states (0 or 1; non-outbreak or outbreak) were assumed. To handle the sporadic extreme values, we included a jump component to filter their negative effects on outbreak detection. Seasonality was handled based on the concept of seasonal random walk.

Our proposed MSJ model is described below:

$$y_t = g\left(Y^{t-1}\right) + z_t \tag{5}$$

$$z_t = \xi_t J_t + x_t \tag{6}$$

$$x_t = a_{0,0} + a_{0,1}s_t + \left(a_{1,0} + a_{1,1}s_t\right)x_{t-1} + \sum_{i=1}^{6} w_i d_{t,i} + \sum_{i=1}^{K} b_i v_{t,i} + e_t \tag{7}$$

$$s_t \in \{0,1\} \tag{8}$$

$$J_t \in \{0,1\} \tag{9}$$

$$p\left(s_t = j \mid s_{t-1} = i\right) = p_{ij} \tag{10}$$

$$e_t \sim N\left(0,\sigma^2\right) \tag{11}$$

$$\xi_t \sim N\left(0,\sigma_a^2\right) \tag{12}$$

$$g\left(Y_{t-1}\right) = med\left\{\bar{y}_{t-m}, \bar{y}_{t-2m}, \bar{y}_{t-3m}\right\} \tag{13}$$

$$\bar{y}_{t-im} = \frac{y_{t-im-3} + y_{t-im-2} + \cdots + y_{t-im+3}}{7} \tag{14}$$

where $Y^{t-1} = (y_1, y_2, \ldots, y_{t-1})$ and $m = 365$. The hidden state variable $s_t = 1$ indicates period t is an outbreak period, 0 otherwise.

Equation 5 filters out the seasonal fluctuation by subtracting the day-of-year expectation from observed time series values. The day-of-year expectation is estimated using the historical values within a day-of-year window in the past three years (Eq. 13-14). The next equation (Eq. 6) further decomposes the residual (z_t) into normal variation (x_t) and a possible jump component. If a jump exists ($J_t = 1$), then ξ_t is the size of the jump. Equation 7 articulates the dynamic behavior during outbreak and non-outbreak periods. The hidden state variable s_t controls the constant term and an autoregressive coefficient. The variables $d_{t,i}$ are day-of-week dummies. The

exogenous variables $v_{t,i}$ are optional controlling factors. Environmental variables such as pollen level and temperature are two possibilities. If necessary, more lagged dependent variables can also be included. For example, we can set $v_{t,1} = x_{t-2}$, $v_{t,2} = x_{t-3}$, ..., $v_{t,6} = x_{t-7}$. As defined in Equation 10, the transition of s_t follows a first-order Markov process.

Compared to conducting outbreak detection using a baseline time series model combined with a statistical surveillance method, our approach provides the following advantages. First, the alert scores ($p(s_t = 1|Y_t)$) of our approach have a clear and intuitive interpretation. Most existing outbreak detecting methods output alert scores that do not have clear meanings. The only way to make sense of the alert scores is to compare the scores with an established threshold. The alert score of our detection algorithm, without reference to any thresholds, can be interpreted as the outbreak probability given available information.

Second, our algorithm provides an estimated outbreak size in addition to outbreak probability. In traditional outbreak detection methods, it is not easy to estimate the outbreak size directly from the alert statistics or estimated parameters. Our method allows the model to recognize different temporal dynamics in different periods. The outbreak size can be calculated directly from the estimated parameters. The information could be valuable for the planning of public health intervention.

Third, the jump component gives our algorithm the ability to separate sporadic extreme values from slow-moving disease outbreaks. The additional information provides flexibility that is valuable for different surveillance needs.

3.1 Changing Dynamics and Outbreak Size

The hidden variable s_t plays an important role in determining the dynamics of x_t. Consider a simplified setting with no day-of-week effect ($w_i = 0$) nor exogenous variables ($b_i = 0$). If we have $s_t = 0$ for all time except $t = t_1$, then the observed value increased by $\Delta_{t1} \equiv a_{0,1} + a_{1,1}y_{t1-1}$ at t_1, ignoring the effect of the noise (e_t). Note that the autoregressive coefficient $a_{1,1}$ also plays an role in determining the magnitude of the increase at time t_1. After this time point, the effect of Δ_{t1} decreases exponentially. The scenario is similar to dropping a group of infected persons in a large community at period t_1 and seeing the disease starting to spread. However, since infected persons recover from the disease quickly, the disease dies out quickly as well.

If $s_t = 1$ for $t = t_1, t_1 + 1, \ldots, t_1 + q$, the effect of increased constant term and autoregressive coefficients accumulates during the outbreak periods until it reaches the new stable level. The new long-term mean can be found by writing x_t as a function of $a_{i,j}$ and e_t only. A simple computation gives $E[x_t|s_t = 1] \equiv \overline{m}_2 = (a_{0,0} + a_{0,1}) / (1 - a_{1,0} - a_{1,1})$. Similarly, the long-term mean of non-outbreak periods is $E[x_t|s_t = 0] \equiv \overline{m}_1 = a_{0,0}/(1 - a_{1,0})$. The outbreak size is the difference between $\overline{m}_2$ and $\overline{m}_1$.

3.2 Model Estimation

Gibbs sampling is used for model estimation. We need to estimate the following sets of coefficients and hidden states: time series coefficients $A = (a_{0,0},a_{0,1},a_{1,0},a_{1,1})$, day-of-week coefficients $W = (w_1,w_2,\ldots,w_6)$, exogenous variable coefficients $B =$

(b_1,b_2,…,b_k), variance of the error term (σ^2), transition probability P = (p_{00},p_{11}), hidden outbreak state S^T = (s_1,s_2,…,s_T), hidden jump state J^T = (J_1,J_2,…,J_T), hidden jump size Ξ^T = (ξ_1,ξ_2,…,ξ_T), and variance of jumps (σ_a^2).

To facilitate the simulation of random variables from the posterior distributions, conjugate priors are used for all parameters. As discussed in the Appendix, all conditional posteriors follow well known statistical distributions and are summarized in Table 1. The dot ($\bullet$) in Table 1 indicates the conditioning on other parameters and hidden states. To increase the efficiency of sampling s_t, the FFBS procedure is used.

TABLE 1
Conditional Posterior Distributions

$(A,W,B)\mid \bullet \sim$ Multivariate Normal
$\sigma^2\mid \bullet \sim$ Inverse Gamma
$\xi_t\mid \bullet \sim$ Normal
$J_t\mid \bullet \sim$ Binomial
$s_t\mid \bullet \sim$ Binomial
$\sigma_a^2\mid \bullet \sim$ Inverse Gamma
$p_{ii}\mid \bullet \sim$ Beta

It should be noted that to avoid the label switching problem, we constraint the parameter sampling results so that $\overline{m}_1 < \overline{m}_2$ is satisfied. If the constraint is violated, (A,W,B) are redrawn until the constraint is satisfied.

3.3 Prospective Outbreak Detection

Given an up-to-date time series, prospective outbreak detection answers the question "What is the probability of having a disease outbreak today?" Letting t denote the current time period, we want to estimate $p(s_t = 1|Y^t)$, where s_t is the hidden outbreak state and Y^t is the vector contains all time series values up to time t. When a new time series value arrives in the next period, the system needs to re-run the model and provide the estimation of $p(s_{t+1}|Y^{t+1})$.

Our preliminary experiments found that direct implementation of the estimation algorithm provides little valuable outbreak information because the algorithm became too sensitive to small changes. The algorithm tried to scrutinize all small changes and tended to over react to those changes. To overcome this difficulty, we developed a regulation technique to desensitize the algorithm so that small, unimportant changes would be ignored.

3.3.1 Desensitization for Prospective Outbreak Detection

The desensitization technique is an extension of the solution for the label switching problem. To make the algorithm ignore small, unimportant changes, we rejected the parameter sampling results that indicated small changes. Specifically, we chose g as the minimal outbreak size that we wanted to detect. We let $a_{0,0}^{(c)}, a_{0,1}^{(c)}, a_{1,0}^{(c)}, a_{1,1}^{(c)}$ be the sampling result of the c-th iteration. We rejected the sampling result if $\overline{m}_1^{(c)} \geq \overline{m}_2^{(c)} - g$. The coefficient g is set to 5% of the time series mean during the training period. Also, the autoregressive coefficient needs to have a value between

-1 and 1 to ensure that the time series is stationary. The desensitization procedure is summarized in Algorithm 1.

Algorithm 1 Desensitization Procedure

repeat

Draw $(A^{(c)},B^{(c)},W^{(c)})$ from $(A,B,W)|\bullet$

$\overline{m}_1^{(c)} \leftarrow a_{0,0}^{(c)}/\left(1-a_{1,0}^{(c)}\right)$

$\overline{m}_2^{(c)} = \left(a_{0,0}^{(c)}+a_{0,1}^{(c)}\right)/\left(1-a_{1,0}^{(c)}-a_{1,1}^{(c)}\right)$

until $\overline{m}_1^{(c)} < \overline{m}_2^{(c)} - g$ and $\left|a_{1,0}\right| < 1$ and $\left|a_{1,0}+a_{1,1}\right| < 1$

return $(A^{(c)},B^{(c)},W^{(c)})$

3.3.2 Prior Distributions

While some parameters of the prior distributions are quite robust to various circumstances, others need to be customized case by case. We applied a simple AR(1) model with day-of-week effect on the training data with seasonality removed. The estimated variance of the error term is used to set up the parameters for the prior of σ^2 and σ_a^2. The estimated day-of-week effects are used to set up the prior of w_i. The prior distributions used in this study are summarized in Table 2.

TABLE 2
Prior Distributions

Parameter	Distribution	Parameter
$\{a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}\}$	Multivariate Normal(M, V)	$M = \{0,0,0.15,0.6\}$, $\{v_{ii}\}_{i=1}^4 = \{400,400,3,3\}$
σ^2	Inverse Gamma	$\alpha = 3, \beta = \text{est. variance} \times (\alpha - 1)$
σ_a^2	Inverse Gamma	$\alpha = 3, \beta = \text{est. variance} \times 5(\alpha - 1)$
$\{w_1, \cdots, w_6\}$	Multivariate Normal(M, V)	M is est. from the training data $\{v_{ii}\}_{i=1}^6 = \{v_a, v_a, v_a, v_a, v_a, v_a\}$; $v_a = \max(100, 5\max(M))$
P_{11}	Beta	a = 2, b = 0.2
P_{22}	Beta	a = 2, b = 0.1

The off-diagnose elements of V is set to zero

3.3.3 Summary of the Estimation Procedure

Given a time series covering period 1 to t_1, our goal is to estimate the outbreak probability of period t_1, together with other relevant parameters and hidden state variables. Using Gibbs sampling for estimation, we need to choose the total number of iteration B and the "burn-in" iteration b. The sampling results between iteration b + 1 and B are then used to compute the outbreak probability (alert score) and the estimates of other parameters. The pseudo code that summarizes the procedure can be found in Algorithm 2. We implemented our approach on R, an open-source statistical software (http://www.r-project.org/).

Algorithm 2 Prospective Outbreak Detection Using the Markov Switching with Jumps (MSJ) Model

for $c = 1$ to B **do**
  $(A^{(c)},B^{(c)},W^{(c)}) \leftarrow$ Desensitization()
  Draw $\sigma^{2(c)}$ from $\sigma^2|\bullet$
  Draw $s_{t_1}^{(c)},s_{t_1-1}^{(c)},\ldots,s_1^{(c)}$ using FFBS
  Draw $J_t^{(c)}$ from $J_t|\bullet$ for $t = 1,2,\ldots,t_1$
  Draw $\xi_t^{(c)}$ from $\xi_t|\bullet$ for $t = 1,2,\ldots,t_1$
  Draw $\sigma_a^{2(c)}$ from $\sigma_a|\bullet$
end for
  $\hat{p}\left(s_{t1}=1\mid Y^t\right)\leftarrow \sum_{c=b+1}^{B} s_{t1}^{(c)}/\left(B-b+1\right)$

APPENDIX A
SELECTED DERIVATION OF THE POSTERIOR DISTRIBUTIONS FOR THE MARKOV SWITCHING WITH JUMPS MODEL

We provide in this appendix the outline for how to derive the conditional posterior distributions. The conditional posterior distributions play a key role in conducting statistical inference. The estimation process iterates to draw random variables from the conditional posteriors in order to construct the joint posterior distribution of parameters and hidden state variables. The discussion is based on the following model:

$$y_t = x_t + \xi_t J_t \tag{15}$$

$$x_t = a_{0,0} + a_{0,1}s_t + \left(a_{1,0} + a_{1,1}s_t\right)x_{t-1} + \sum_{i=1}^{6} d_{t,i}w_i + e_t \tag{16}$$

$$s_t \in \{0,1\} \tag{17}$$

$$J_t \in \{0,1\} \tag{18}$$

$$p\left(s_t = j \mid s_{t-1} = i\right) = p_{ij} \tag{19}$$

$$p\left(J_t = 1\right) = q_j \tag{20}$$

$$e_t \sim N\left(0,\sigma^2\right) \tag{21}$$

$$\xi_t \sim N\left(0,\sigma_a^2\right) \tag{22}$$

Note that this model is slightly different from the one used in our study. The main difference is that we assume that the time series has been "preprocessed" to remove seasonality. So the y_t here is equivalent to z_t in Equation 5. Also, without loss of generality, b_i is assumed to be zero. We applied Bayesian inference techniques in this study [73]. Specifically, Gibbs sampling was used. Our basic model has the following state variables: $S^T = (s_1,s_2,...,s_T)$, $J^T = (J_1,J_2,...,J_T)$, and $\Xi^T = (\xi_1,\xi_2,...,\xi_T)$. Although $X^T = (x_1,x_2,...,x_T)$ is not observed either, the values are fully determined if both J^T and Ξ^T are known. The coefficients to be estimated are denoted by $\Theta = (a_{0,0},a_{0,1},a_{1,0},a_{1,1},p_{11},p_{22},q_1,\sigma^2,\sigma_a^2,w_1,...,w_6)$.

Using the Gibbs sampling technique, we approximate the posterior distribution of parameters and hidden state variables, $p(\Xi^T, J^T, S^T, \Theta | Y^T)$, by iteratively drawing random variables from the following conditional distributions:

$$p\left(\Xi^T \mid J^T, S^T, \Theta, Y^T\right)$$
$$p\left(J^T \mid \Xi^T, S^T, \Theta, Y^T\right)$$
$$p\left(S^T \mid \Xi^T, J^T, \Theta, Y^T\right)$$
$$p\left(\Theta \mid \Xi^T, J^T, S^T, Y^T\right)$$

The following is the derivation of the conditional posterior distributions.

A.0.1 Drawing from $p(\Xi^T \mid J^T, S^T, \Theta, Y^T)$

To draw Ξ^T from its conditional posterior, we iterate through each period and draw ξ_t given $\xi_{-t} = \{\xi_1, \ldots, \xi_{t-1}, \xi_{t+1}, \ldots, \xi_T\}$ and other random variables. Consider the jump size ξ_t at period t when the corresponding indicator variable $J_t = 1$.

$$\begin{aligned}
& p\left(\xi_t \mid \xi_{-t}, J^T, Y^T, S^T, \Theta\right) \\
\propto\ & p\left(Y^T, \xi_{-t}, J^T \mid \xi_t, S^T, \Theta\right) p\left(\xi_t \mid S^T, \Theta\right) \\
\propto\ & p\left(Y^T \mid \xi_{-t}, J^T, \xi_t, S^T, \Theta\right) p\left(\xi_{-t}, J^T \mid \xi_t, S^T, \Theta\right) p\left(\xi_t \mid S^T, \Theta\right) \\
\propto\ & p\left(y_{t+1} \mid y_t, \Xi_T, J^T, S^T, \Theta\right) p\left(y_t \mid y_{t-1}, \Xi_T, J^T, S^T, \Theta\right) p(\xi) \\
\equiv\ & p\left(y_{t+1} \mid y_t\right) p\left(y_t \mid y_{t-1}\right) p(\xi)
\end{aligned}$$

Note that $p(\xi_t | S^T, \Theta) = p(\xi_t)$ and $p(\xi_{-t}, J^T | \xi_t, S^T, \Theta) = p(\xi_{-t}) p(J^T)$ by definition. To make the equations easier to read, we suppressed the conditioning on Ξ_T, J^T, S^T, Θ at (23) and in the following discussion.

Since

$$p\left(y_{t+1} \mid y_t\right) \propto \frac{-1}{\sigma^2}\left[y_{t+1} - J_{t+1}\xi_{t+1} - \left(a_{0,0} + a_{0,1}s_{t+1}\right) - \left(a_{1,0} + a_{1,1}s_{t+1}\right)\left(y_t - J_t\xi_t\right)\right]^2 \quad (24)$$

$$p\left(y_t \mid y_{t-1}\right) \propto \frac{-1}{\sigma^2}\left[y_t - J_t\xi_t - \left(a_{0,0} + a_{0,1}s_t\right) - \left(a_{1,0} + a_{1,1}s_t\right)\left(y_{t-1} - J_{t-1}\xi_{t-1}\right)\right]^2 \quad (25)$$

$$p\left(\xi\right) \propto \frac{-1}{\sigma_a^2}\xi_t^2 \quad (26)$$

Substituting (24) – (26) back to (23) and complete square with respect to ξ_t, we get the conditional posterior distribution of ξ_t:

$$\xi_t \mid J_t = 1 \sim N\left(m_t, v_t\right) \quad (27)$$
$$\xi_t \mid J_t = 0 \sim N\left(0, \sigma_a^2\right) \quad (28)$$

where

$$
\begin{aligned}
m_t &= \frac{\sigma_a^2\left(z_t - \phi_{t+1} z_{t+1}\right)}{\sigma_a^2\left(1+\phi_{t+1}^2\right)+\sigma^2} \\
v_t &= \frac{\sigma^2\sigma_a^2}{\sigma^2 + \left(1+\phi_{t+1}^2\right)\sigma_a^2} \\
z_t &= y_t - \alpha_t - \phi_t\left(y_{t-1} - \xi_{t-1} J_{t-1}\right) \\
z_{t+1} &= y_{t+1} - \xi_{t+1} J_{t+1} - \alpha_{t+1} - \phi_{t+1} y_t \\
\alpha_t &= a_{0,0} + a_{0,1} s_t \\
\phi_t &= a_{1,0} + a_{1,1} s_t
\end{aligned}
$$

A.0.2 Drawing from $p(J^T \mid \Xi^T, S^T, \Theta, Y^T)$

The posterior of J_t is

$$
\begin{aligned}
& p\left(J_t \mid \Xi_T, Y^T, J_{-t}, S^T, \Theta\right) \\
\propto\ & p\left(Y^T \mid J_t, J_{-t}, \Xi_t, S^T, \Theta\right) p\left(J_t \mid \Xi_t, S^T, \Theta\right) \\
\propto\ & p\left(y_{t+1} \mid y_t, \Xi_T, J^T, S^T, \Theta\right) p\left(y_t \mid y_{t-1}, \Xi_T, J^T, S^T, \Theta\right) p\left(J_t\right) \\
\equiv\ & p\left(y_{t+1} \mid y_t\right) p\left(y_t \mid y_{t-1}\right) p\left(J_t\right)
\end{aligned}
\tag{29}
$$

The posterior probability of $J_t = 1$ can be calculated by considering the odd ratio

$$
\frac{p\left(J_t = 1 \mid \bullet\right)}{p\left(J_t = 0 \mid \bullet\right)} \tag{30}
$$

A.0.3 Drawing from $p(S^T \mid \Xi^T, J^T, \Theta, Y^T)$

Since $x_t = y_t - \xi_t J_t$, the conditional posterior $p(S^T \mid \Xi^T, J^T, \Theta, Y^T)$ can be written as $p(S^T \mid X^T, \Theta)$. Multi-move Gibbs sampling is used to draw S^T from its posterior. To achieve this, the first step is to calculate the filtered state probabilities, i.e., $p(s_t = l \mid X^t)$, $l \in \{0,1\}$. The calculation can be divided into three steps:

(1) One-step ahead prediction of s_t:

$$
p\left(s_t = l \mid X^{t-1}\right) = \sum_{k=0}^{1} p_{kl}\, p\left(s_{t-1} = l \mid X^{t-1}\right) \tag{31}
$$

(2) Filtering for s_t

$$
p\left(s_t = l \mid X^t\right) = \frac{p\left(x_t \mid s_t = l, X^{t-1}\right) p\left(s_t = l \mid X^{t-1}\right)}{p\left(x_t \mid X^{t-1}\right)} \tag{32}
$$

where

$$
p\left(x_t \mid X^{t-1}\right) = \sum_{k=0}^{1} p\left(x_t \mid S_t = k, X^{t-1}\right) p\left(s_t = k \mid x^{t-1}\right) \tag{33}
$$

The smoothed probability $p(S^T \mid X^T, \Theta)$ can be calculated as follows:

$$p\left(S_t = l \mid X^T\right) = \sum_{k=0}^{1} \frac{p_{lk}\, p\left(s_t = l \mid X^t\right) p\left(s_{t+1} = k \mid X^T\right)}{\sum_{j=0}^{1} p_{jk}\, p\left(S_t = j \mid X^t\right)} \tag{34}$$

The multi-move Gibbs sampling makes use of the following expansion for S^T:

$$\begin{aligned} & p\left(S^T \mid X^T\right) \\ = \; & p\left(s_T \mid X^T\right) p\left(s_{T-1} \mid s_T, X^{T-1}\right) p\left(s_{T-2} \mid s_{T-1}, X^{T-2}\right) \cdots p\left(s_1 \mid s_2, x_1\right) \\ = \; & p\left(s_T \mid X^T\right) \prod_{t=1}^{T-1} p\left(s_t \mid s_{t+1}, X^t\right) \end{aligned}$$

where

$$p\left(s_t \mid X^t, s_{t+1}\right) \propto p\left(s_{t+1} \mid s_t\right) p\left(s_t \mid X^t\right) \tag{35}$$

A.0.4 Drawing from $p(\Theta^T \mid \Xi^T, J^T, S^T, Y^T)$

Given state variable S^T and jump variables J^T, Ξ^T, the posterior distribution of $a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}, w_i, \sigma^2$ follows from the standard Bayesian regression model.

Specifically, let $m_t = \{1, s_t, x_{t-1}, s_t x_{t-1}, d_{t,1}, \ldots, d_{t6}\}$ be a row vector, then $M^T = \{m_1' m_2' \ldots m_T'\}'$ is a matrix with T rows. Then the posterior of $\beta = \{a_{0,0}, a_{0,1}, a_{1,0}, a_{1,1}, w_1, w_2, \ldots w_6\}$ follows a normal distribution

$$\beta \sim N\left(u_\beta, v_\beta\right) \tag{36}$$

$$u_\beta = v_\beta \left(\frac{M'X}{\sigma^2} + v_0^{-1} \beta_0 \right) \tag{37}$$

$$v_\beta^{-1} = \frac{M'M}{\sigma^2} + v_0^{-1} \tag{38}$$

The posterior distribution of σ^2 follows Inverse Gamma distribution

$$\sigma^2 \sim IG\left(v_g, \lambda_g\right) \tag{39}$$

$$v_g = v_{g0} + (T)/2 \tag{40}$$

$$\lambda_g = \lambda_{g0} \left(U^T{}' U^T \right) \Big/ 2 \tag{41}$$

$$U^T = X^T - M^T \beta \tag{42}$$

The posteriors of p_{00} and p_{11} are

$$p_{00} \mid S^T \sim beta\left(u_{00} + n_{00}, u_{01} + n_{01}\right) \tag{43}$$

$$p_{11} \mid S^T \sim beta\left(u_{11} + n_{11}, u_{10} + n_{10}\right) \tag{44}$$

where n_{ij} refers to the count of transitions from state i to j, which can be calculated directly from S^T. u_{ij} refers to the parameters of the prior distributions for p_{00} and p_{11}.

The posterior of q_1 is

$$q_1 \mid S^T \sim beta(v_1 + n_{v1}, v_0 + n_{v0}) \tag{45}$$

where n_{v1} is the count of $J_t = 1$ and n_{v0} is the count of $J_t = 0$. v_1 and v_0 are the parameters of the prior distribution of q_1.

REFERENCES

[1] P.-H. Hu, D. Zeng, H. Chen, C. Larson, W. Chang, C. Tseng, and J. Ma, "System for infectious disease information sharing and analysis: Design and evaluation," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 11, no. 4, pp. 483–492, July 2007.

[2] S. Niiranen, J. Yli-Hietanen, and L. Nathanson, "Toward reflective management of emergency department chief complaint information," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 12, no. 6, pp. 763–767, Nov. 2008.

[3] W. W. Chapman, J. N. Dowling, and M. M. Wagner, "Generating a reliable reference standard set for syndromic case classification," *Journal of the American Medical Informatics Association*, vol. 12, no. 6, pp. 618–629, 2005.

[4] W. W. Chapman, L. M. Christensen, M. M. Wagner, P. J. Haug, O. Ivanov, J. N. Dowling, and R. T. Olszewski, "Classifying free-text triage chief complaints into syndromic categories with natural language processing," *Artificial Intelligence in Medicine*, vol. 33, no. 1, pp. 31–40, 2005.

[5] O. Ivanov, M. M. Wagner, W. W. Chapman, and R. T. Olszewski, "Accuracy of three classifiers of acute gastrointestinal syndrome for syndromic surveillance," in *Proceedings of the AMIA Symposium*, 2002, pp. 345–349.

[6] J. U. Espino and M. Wagner, "The accuracy of ICD-9 coded chief complaints for detection of acute respiratory illness," in *Proceedings of the AMIA Annual Symposium*, 2001, pp. 164–168.

[7] P. Yan, H. Chen, and D. Zeng, "Syndromic surveillance systems," *Annual Review of Information Science and Technology*, vol. 42, 2007.

[8] J. Espino, M. Wagner, F. Tsui, H. Su, R. Olszewski, Z. Lie, W. Chapman, X. Zeng, L. Ma, Z. Lu, and J. Dara, "The RODS Open Source Project: removing a barrier to syndromic surveillance," *Stud Health Technol Inform*, vol. 107, pp. 1192–1196, 2004.

[9] J. Lombardo, H. Burkom, E. Elbert, S. Magruder, S. H. Lewis, W. Loschen, J. Sari, C. Sniegoski, R. Wojcik, and J. Pavlin, "A system overview of the electronic surveillance system for early notification of community-based epidemics (ESSENCE II)," *Journal of Urban Health*, vol. 80, no. 2, pp. i32–i42, 2003.

[10] B. M. Lawson, E. C. Fitzhugh, S. P. Hall, C. Franklin, L. C. Hutwagner, G. M. Seeman, and A. S. Craig, "Multifaceted syndromic surveillance in a public health department using the early aberration reporting system," *Journal of Public Health Management and Practice*, vol. 11, no. 4, pp. 274–281, 2005.

[11] K. D. Mandl, M. Overhage, M. Wagner, W. Lober, P. Sebastiani, F. Mostashari, J. Pavlin, P. H. Gesteland, T. Treadwell, E. Koski, L. Hutwagner, D. L. Buckeridge, R. D. Aller, and S. Grannis, "Implementing syndromic surveillance: a practical guide informed by the early experience," *Journal of the American Medical Informatics Association*, vol. 11, no. 2, pp. 141–150, 2004.

[12] L. Hutwagner, W. Thompson, G. M. Seeman, and T. Treadwell, "The bioterrorism preparedness and response Early Aberration Reporting System (EARS)," *Journal of Urban Health*, vol. 80, no. 2, Supplement 1, pp. i89–i96, 2003.

[13] B. Y. Reis and K. D. Mandl, "Integrating syndromic surveillance data across multiple locations: Effects on outbreak detection performance," in *AMIA 2003 Symposium Proceedings*, 2003, pp. 549–553.

[14] B. Y. Reis, M. Pagano, and K. D. Mandl, "Using temporal context to improve biosurveillance," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 1961–1965, Feb 2003.

[15] B. Y. Reis and K. D. Mandl, "Time series modeling for syndromic surveillance," *BMC Med Inform Decis Mak*, vol. 3, p. 2, Jan 2003.

[16] J. Takeuchi and K. Yamanishi, "A unifying framework for detecting outliers and change points from time series," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 4, pp. 482–492, 2006.

[17] W. A. Shewhart, *Statistical method from the viewpoint of quality control*. Washington, The Graduate School, The Department of Agriculture, 1939.

[18] D. C. Montgomery, *Introduction to statistical quality control*, 5th ed. Wiley, New York, 2005.

[19] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, jun 1954.

[20] CDC, "Increased antiviral medication sales before the 2005-06 influenza season–New York City," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 55, pp. 277–279, Mar 2006.
[21] D. L. Buckeridge, P. Switzer, D. Owens, D. Siegrist, J. Pavlin, and M. Musen, "An evaluation model for syndromic surveillance: assessing the performance of a temporal algorithm," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 54 Suppl, pp. 109–115, Aug 2005.
[22] M. P. Clements and D. F. Hendry, *Handbook of Economic Forecasting*. Elsevier, 2006, vol. 1, ch. Forecasting with breaks, pp. 605 – 657.
[23] C.-S. J. Chu, M. Stinchcombe, and H. White, "Monitoring structural change," *Econometrica*, vol. 64, no. 5, pp. 1045–1065, 1996.
[24] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, no. 2, pp. 357–84, March 1989.
[25] G. Box and G. Jenkins, *Time series analysis: Forecasting and control, San Francisco: Holden-Day*, 1970.
[26] W. H. Greene, *Econometric Analysis*. Prentice Hall, 2000.
[27] H. Akaike, "Statistical predictor identification," *Annals of the Institute of Statistical Mathematics*, vol. 22, pp. 203–217, 1970.
[28] ——, "Information theory and an extension of the likelihood principle," in *Proceedings of the Second International Symposium of Information Theory*, B. N. Perov and F. Csaki, Eds., Akademiai Kiado, Budapest, 1973.
[29] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
[30] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[31] H. White, *Approximate Nonlinear Forecasting Methods*, ser. Handbook of Economic Forecasting. Elsevier, January 2006, vol. 1, ch. 9, pp. 459–512.
[32] J. Shao, "An asymptotic theory for linear model selection," *Statistica Sinica*, vol. 7, pp. 221–264, 1997.
[33] M. L. Jackson, A. Baer, I. Painter, and J. Duchin, "A simulation study comparing aberration detection algorithms for syndromic surveillance," *BMC Med Inform Decis Mak*, vol. 7, p. 6, 2007.
[34] S. C. Wieland, J. S. Brownstein, B. Berger, and K. D. Mandl, "Automated real time constant-specificity surveillance for disease outbreaks," *BMC Medical Informatics and Decision Making*, vol. 7, no. 15, 2007.
[35] J. Zhang, F.-C. Tsui, and M. M. W. andWilliam R. Hogan, "Detection of outbreaks from time series data using wavelet transform," in *Proc AMIA Symp*, 2003.
[36] R. Serfling, "Methods for current statistical analysis of excess pneumonia- influenza deaths," *Public Health Reports*, vol. 78, pp. 494–506, 1963.
[37] J. C. Brillman, T. Burr, D. Forslund, E. Joyce, R. Picard, and E. Umland, "Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance," *BMC Med Inform Decis Mak*, vol. 5, p. 4, 2005.
[38] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, 2004.
[39] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, no. 3, pp. 324–342, 1960.
[40] H. S. Burkom, S. P. Murphy, and G. Shmueli, "Automated time series forecasting for biosurveillance," *Stat Med*, vol. 26, pp. 4202–4218, Sep 2007.
[41] J. Hamilton, *Time Series Analysis*. Princeton, 1994.
[42] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability and Its Applications*, vol. 8, pp. 22–46, 1963.
[43] S. W. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, pp. 411–430, 1966.
[44] M. Frisen and J. De Mare, "Optimal surveillance," *Biometrika*, vol. 78, no. 2, pp. 271–280, 1991.
[45] C. Sonesson and D. Book, "Review and discussion of prospective statistical surveillance in public health," *Journal of the Royal Statistical Society, Series A*, vol. 166, no. 1, pp. 5–21, 2003.
[46] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, dec 1986.
[47] M. Frisen, "Statistical surveillance. optimality and methods," *International Statistical Review*, vol. 71, no. 2, pp. 403–434, 2003.
[48] S. Chandrasekaran, J. R. English, and R. L. Disney, "Modeling and analysis of ewma control schemes with varance-adjusted control limits," *IIE Transactions*, vol. 27, pp. 282–290, 1995.
[49] S. H. Steiner, "Ewma control charts with time-varying control limits and fast initial response," *Journal of Quality Technology*, vol. 31, no. 1, pp. 75–86, 1999.
[50] C. Sonesson, "Evaluations of some exponentially weighted moving average methods," *Journal of Applied Statistics*, vol. 30, no. 10, pp. 1115–1133, 2003.
[51] H. Burkom, *Disease Surveil1ance:A Public Health Informatics Approach*. John Wiley & Sons,, 2007, ch. Alerting Algorithms for Biosurveillance, pp. 143–192.
[52] C.-J. Kim and C. R. Nelson, *State-space models with regime switching*. MIT Press, Cambridge, 1999.

[53] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state markov chains," *Annals of Math. Statistics*, vol. 37, pp. 1554–1563, 1966.
[54] L. E. Baum and J. A. Egon, "An inequality with applications to statistical estimation for probabilistic functions of a markov process and to a model for ecology," *Bull. Amer. Meteorology Soc.*, vol. 73, pp. 360–363, 1967.
[55] Y. L. Strat and F. Carrat, "Monitoring epidemiologic surveillance data using hidden markov models," *Statistics in Medicine*, vol. 18, pp. 3463–3478, 1999.
[56] M. Dahlquist and S. F. Gray, "Regime-switching and interest rates in the european monetary system," *Journal of International Economics*, vol. 50, no. 2, pp. 399–419, April 2000.
[57] S. L. Scott, "Bayesian methods for hidden markov models: recursive computing in the 21st century," *Journal of the American Statistical Association*, vol. 97, pp. 337–351, 2002.
[58] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
[59] C. A. Popescu and Y. S. Wong, "Nested Monte Carlo EM algorithm for switching state-space models," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 12, pp. 1653–1663, 2005.
[60] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 5, pp. 631–645, 2007.
[61] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, nov 1995.
[62] J. H. Albert and S. Chib, "Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts," *Journal of Business & Economic Statistics*, vol. 11, no. 1, pp. 1–15, jan 1993.
[63] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrika*, vol. 81, no. 3, pp. 541–553, aug 1994.
[64] D. Madigan, *Spatial and Syndromic Surveillance for Public Health*. John Wiley & Sons, 2005, ch. Bayesian Data Mining for Health Surveillance, pp. 203–221.
[65] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 2, pp. 192–236, 1974.
[66] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.
[67] ISDS, "My algorithm can out-detect your algorithm: Biosurveillance using time series data," International Society for Disease Surveillance, Tech. Rep., 2008, https://wiki.cirg.washington.edu/pub/bin/view/Isds/TechnicalContest ; accessed Nov. 23, 2008.
[68] L. M. Wein, D. L. Craft, and E. H. Kaplan, "Emergency response to an anthrax attack," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, pp. 4346–4351, Apr 2003.
[69] J. A. Jernigan, D. S. Stephens, D. A. Ashford, C. Omenaca, M. S. Topiel, M. Galbraith, M. Tapper, T. L. Fisk, S. Zaki, T. Popovic, R. F. Meyer, C. P. Quinn, S. A. Harper, S. K. Fridkin, J. J. Sejvar, C. W. Shepard, M. McConnell, J. Guarner, W. J. Shieh, J. Malecki, J. L. Gerberding, J. M. Hughes, and B. A. Perkins, "Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States," *Emerging Infect. Dis.*, vol. 7, pp. 933–944, 2001.
[70] T. Burr, T. Graves, R. Klamann, S. Michalak, R. Picard, and N. Hengartner, "Accounting for seasonal patterns in syndromic surveillance data for outbreak detection," *BMC Medical Informatics and Decision Making*, vol. 6, no. 40, 2006.
[71] H. Rolka, H. Burkom, G. F. Cooper, M. Kulldorff, D. Madigan, and W.-K. Wong, "Issues in applied statistics for public health bioterrorism surveillance using multiple data stream: research needs," *Statistics in Medicine*, vol. 26, pp. 1834–1856, 2007.
[72] H.-M. Lu, D. Zeng, L. Trujillo, K. Komatsu, and H. Chen, "Ontology-enhanced automatic chief complaint classification for syndromic surveillance." *J Biomed Inform*, vol. 41, no. 2, pp. 340–356, Apr 2008.
[73] R. Chang, M. Stetter, and W. Brauer, "Quantitative inference by qualitative semantic knowledge mining with bayesian model averaging," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 12, pp. 1587–1600, 2008.