A Three-State Recursive Sequential Bayesian Algorithm for Biosurveillance

K. D. Zamba [a] Panagiotis Tsiamyrtzis [b] Douglas M. Hawkins [c]

[a] Department of Biostatistics, The University of Iowa, 200 Hawkins Drive, C22M GH Iowa City, IA 52242

[b] Department of Statistics, Athens University of Economics and Business; 76 Patission Str 10434 Athens Greece

[c] School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street Minneapolis, MN 55455

Abstract

A serial signal detection algorithm is developed to monitor pre-diagnosis and medical diagnosis data pertaining to biosurveillance. The algorithm is three-state sequential, based on Bayesian thinking. It accounts for non-stationarity, irregularity, seasonality, and captures an epidemic serial structural details. At stage n, a trichotomous variable governing the states of an epidemic is defined, and a prior distribution for time-indexed serial readings is set. The technicality consists of finding a posterior state probability based on the observed data history, using the posterior as a prior distribution for stage $n+1$ and sequentially monitoring surges in posterior state probabilities. A sensitivity analysis for validation is conducted and analytical formulas for the predictive distribution are supplied for error management purposes. The method is applied to syndromic surveillance data gathered in the United States (U.S.) District of Columbia metropolitan area.

Keywords: Bayesian Sequential Update, Dynamic Control, Syndromic Surveillance.

1 Introduction

The problem that led to developing the method herein came from medical arena. Data pertaining to biosurveillance are used in the U.S. and in some countries around the globe to

assess data-driven evidence for natural epidemics or intentional release of biological agents with operations similar to natural epidemics. With the view of responding to the serious danger of bioterrorism, there has been a shift in disease surveillance from the classical retrospective disease chart review, to a prospective and real-time early disease detection. In syndromic surveillance for example, investigators monitor non-specific clinical information that may indicate bioterrorism associated disease before specific diagnoses are made. Medical related data such as emergency room observations, over-the-counter sales, veterinary data, medical and public health information, are used as investigational tools to screen out information about outbreaks and augment early reports to sentinels. Relevant publications describing the need to integrate medical related data into real-time surveillance systems are Arnon et al. (2001); Dennis et al. (2001); Henderson et al. (1999); Inglesby et al. (1999, 2000). An overview of the use of syndromic surveillance can be found in Buehler et al. (2003); overviews and examples of syndromic surveillance systems are found in Green and Kaufman (2002); and description of the steps of disease outbreak investigation and syndromic surveillance are described in Pavlin (2003).

Statistical process control (SPC) tools have long time been central parts of classical disease surveillance and laboratory based outbreak detections. It is to be noted that the U.S. Center for Disease Control (CDC) has routinely applied cumulative sum (Cusum) techniques to laboratory-based data for outbreak detection; see for example Hutwagner et al. (1997), Stern and Lightfoot (1999). These tools though, need extra maintenance if they were to migrate to the field of modern biosurveillance. Biosurveillance data are not like industrial and laboratory data for which traditional SPC methods are developed; they fail to meet most underlying assumptions of SPC. For example, the main parameters associated with incidence in syndromic surveillance may be unknown, or only partially known, or unstable and may change a course from year to year. In addition, incidence parameters may jump at any time and in any direction (upward/downward or stable). It has been recognized that for some infectious diseases with seasonal trend such as influenza, each season is likely to generate its specific data pattern; a new data process to be studied *de novo*.

A critical requirement for biosurveillance algorithms is fast detection. This requirement is quickly clouded by a background of variability and structural changes depicted on biosurveillance data. As common feature, these data substantially violate the various statistical underpinnings of standard SPC techniques (we refer a reader less familiar with SPC methods in medical surveillance to Woodall, 2006; Zamba et al., 2008; Shmueli and Burkom, 2010 for works and in-depth discussions on assumptions and their violations). To explain, as back-

ground variability and structural changes affect these data, the monitoring scheme used for signal detection yields more false signals than anticipated; thus resulting in low sensitivity and low specificity; Stoto et al. (2004). In addition to variability and structural changes, historical data gatherings do not provide consistent estimates of the true parameters to monitor due to irregular pattern of the underlying epidemics. Non-stationarity, unpredictability, irregularity and parameter instability are common features of biosurveillance data.

The context and requirement of biosurveillance (timeliness and fast detection) would demand individual observation SPC methods if underlying assumptions were met. Cusums, exponentially weighted moving averages (Ewma), and change point charts might seem to be first blush natural candidates. These charts too, are limited. For example, Cusums are governed by statistical assumptions such as independent readings, normal distribution with known mean, known standard deviation, and a requirement of extensive/stable historical data set (HDS); see Hawkins and Olwell, 1998. These statistical underpinnings of Cusums make them lesser suitable candidates for monitoring real-time biosurveillance data. In addition, Cusums are better known for responding to step (not gradual) changes. Ewma requires advance knowledge and stability of the process parameters. Change-point methods, although less demanding in their calibration needs and better suited for handling individual reading and start-up processes, rely on independent and identically distributed normal assumption along with an assumption of change from a constant set of distributional parameters to another constant set (see for example Hawkins et al., 2003; Hawkins and Zamba, 2005a, 2005b). Another SPC method commonly used, is to fit a model to data readings and chart the residuals. This method is known as residual charts and is frequently applied to autocorrelated data. Residual charts are not suited for monitoring individual data readings. Their performance is not stellar, and have the disadvantage of filtering observations through a model before charting (see Zhang, 1997; Winkel and Zhang, 2007). Residual charts can go as far as the model which yields the residuals.

Our goal is to develop a fast signal detection method to monitor these types of data pertaining to biosurveillance. We propose a three-state recursive and sequentially updated method. It is an individual data monitoring scheme (case). The approach is Bayesian and sets a prior distribution for the serial evolution of the study variable. The method then finds a posterior based on observed data, assesses the state of an epidemic and uses this posterior as a prior distribution for the next stage. The Bayesian approach appears to be quite appealing in modeling these data because at best, the historical information although imperfect and short, can be used to elicit an appropriate prior. Furthermore, the need to draw inference in

an online fashion even in a presence of a short run, along with the possibility of deriving the predictive distribution for the future observable(s) make the Bayesian approach a favorable alternative to traditional SPC methods.

The manuscript is organized as follow: Section 2 defines our proposed three-state model and the algebraic formulas governing the states at each stage. Section 3 elaborates on the attractiveness of our model, its novelty and its reproducibility. Section 4 outlines our simulation results, the sensitivity analysis, and applies our approach to real-time data gathered in the U.S. We close our paper with sections on comparative study, discussion and technical appendix.

2 Proposed Three-State Bayesian Statistical Modeling

We assume that observations are gathered sequentially and define a time indexed series $Y_n, n = 1, 2, \ldots$; where n may be hours, days, weeks, months, ..., depending on the importance of the characteristic being monitored. The observed Y_n may be actual case counts or percentage of some activity observed, standardized observed cases, or any aggregation of observed data over a specific hospital, location county, or state. We work under the assumption that readings are large enough for normality to hold.

At time $n = 0$, before any monitoring starts, we set a prior distribution for the response variable; we denote this by Y_0:

$$Y_0 \sim N(\zeta, \sigma_0^2),$$

where ζ and σ_0^2 are the initial prior mean and variance, assumed known. The prior information, although imperfect, may come from an educated guess, from previous HDS or from expert's opinion.

Moving between successive stages $n-1$ and n, we model Y_n using a first order autoregressive AR(1) model:

$$Y_n = \lambda Y_{n-1} + u_n, \tag{1}$$

where $\lambda \in (0, 1)$ is a discount factor, assumed to be known, and u_n is a random error displacement. Observe that for $\lambda = 0$, the model becomes the random batch effect model while for $\lambda = 1$ the model is a random walk model. The AR representation provides a smooth transition between successive stages, relaxing the independence assumption governing most standard SPC techniques. The three-state approach comes as follow: Y_n may be affected by three mutually exclusive states; *flat, rise* and *decay*.

- *Flat state:* Y_n does not change much from Y_{n-1}

- *Rise state:* Y_n experiences positive jumps/shocks with respect to Y_{n-1}
- *Decay state:* Y_n experiences negative jumps/shocks with respect to Y_{n-1}.

We model these three states using a trichotomous random variable E_n. Specifically, at each stage n one has:

$$E_n = \left\{ \begin{array}{lll} 0 & \text{with prob.} & \pi_0^{(n-1)} \\ -1 & \text{with prob.} & \pi_-^{(n-1)} \\ +1 & \text{with prob.} & \pi_+^{(n-1)} \end{array} \right\}, \tag{2}$$

with $0, -1$ and $+1$ indicating *flat*, *decay* and *rise* states respectively; each of which occurs with some a-priori probability $\left(\pi_0^{(n-1)} + \pi_-^{(n-1)} + \pi_+^{(n-1)} = 1\right)$.

Each realization of E_n corresponds to either a jump (for the $E_n = \pm 1$) or a no-jump ($E_n = 0$) scenario. For the cases where a jump occurs (i.e. *rise* or *decay* states) we consider the size of the jump to be a random variable with directionality given by the sign of E_n (the direction is upward for the *rise* and downward for the *decay* states). If δ_n denotes the size of the jump at time n, we assume:

$$\delta_n \sim N(\Delta, \tau^2),$$

where Δ is the expected magnitude of the jump, and τ^2 is the uncertainty regarding the prior mean estimate Δ. In this setting we have selected the same expected magnitude for both the *rise* and the *decay* because of the hallmark of our illustrative epidemics under study. One may choose different magnitudes, Δ_1 for rise and Δ_2 for decay, based on a specific epidemic or upon observing the shape of an epidemic curve using a small but relevant HDS.

Brief, E_n is a trichotomous random variable associated with the directionality of Y_n and governs the states, while δ_n is associated with the size of the random shock. Both these entities are embedded in the random error displacement u_n of Equation (1) as follows:

$$u_n | (E_n, \delta_n) \sim N(E_n \delta_n, \sigma^2);$$

yielding an unconditional distribution of random error u_n of

$$u_n \sim \left\{ \begin{array}{lll} N(\,0,\ \sigma^2) & \text{with prob.} & \pi_0^{(n-1)} \\ N(-\Delta,\ \sigma^2 + \tau^2) & \text{with prob.} & \pi_-^{(n-1)} \\ N(+\Delta,\ \sigma^2 + \tau^2) & \text{with prob.} & \pi_+^{(n-1)} \end{array} \right\}.$$

Thus, the random error term in Equation (1) corresponds to a mixture of three normal distributions, each of which is specifically mapped onto one of the three states of an epidemic under study at each stage. Our method for monitoring the series is a sequentially updated scheme based on monitoring two out of three states posterior probabilities:

- At stage n, get the observation Y_n and obtain the posterior $E_n|Y_n$.
- Draw a decision regarding the state of the epidemics, based on the posterior distribution and obtain the predictive distribution.
- Continue monitoring if the posterior distribution is within the 'range of acceptability' (point to which we will return).
- Use the posterior distribution at stage n as a prior distribution for stage $n+1$ (sequential update).

In Equation (2), the probabilities $\pi_0^{(n-1)}$, $\pi_-^{(n-1)}$ and $\pi_+^{(n-1)}$ require some attention. These respectively refer to the posterior probabilities of *flat*, *decay* and *rise* states of an epidemic at time point $n-1$. For this scheme to be able to run, one needs to set some initial prior probabilities $\pi_0^{(0)}$, $\pi_-^{(0)}$ and $\pi_+^{(0)}$ which reflect prior beliefs regarding the states of an epidemic at time 0. In the event that no such prior belief exists, one can simply choose a non-informative prior such as $\pi_0^{(0)} = \pi_-^{(0)} = \pi_+^{(0)} = 1/3$.

The following theorem is the basis for our recursive monitoring:

Theorem 1 *The unconditional distribution of Y_n at each time n is a mixture of 3^n Normal distributions.*

$$f(Y_n) = \sum_{i=0}^{3^n-1} w_i^{(n)} N\left(\mu_i^{(n)}, \left(\sigma_i^{(n)}\right)^2\right).$$

The weights, means and variances, for $n \geq 1$ and for $i = 0, 1, \ldots, 3^{n-1}-1$ obey the recursions:

$$\begin{array}{lllllllll}
w_{3i}^{(n)} & = & w_i^{(n-1)}\pi_0^{(n-1)} & \mu_{3i}^{(n)} & = & \lambda\mu_i^{(n-1)} & \left(\sigma_{3i}^{(n)}\right)^2 & = & \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2 \\
w_{3i+1}^{(n)} & = & w_i^{(n-1)}\pi_-^{(n-1)} & \mu_{3i+1}^{(n)} & = & \lambda\mu_i^{(n-1)} - \Delta & \left(\sigma_{3i+1}^{(n)}\right)^2 & = & \tau^2 + \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2 \\
w_{3i+2}^{(n)} & = & w_i^{(n-1)}\pi_+^{(n-1)} & \mu_{3i+2}^{(n)} & = & \lambda\mu_i^{(n-1)} + \Delta & \left(\sigma_{3i+2}^{(n)}\right)^2 & = & \tau^2 + \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2,
\end{array}$$

with initial values $w_i^{(0)} = 1$, $\mu_i^{(0)} = \zeta$ and $\sigma_i^{(0)} = \sigma_0$; the posterior probabilities are given by:

$$P(E_n = 0|Y_n = y_n) = \frac{\displaystyle\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_0^{(n-1)}}{\sqrt{2\pi\left[\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\lambda\mu_i^{(n-1)}\right)^2}{2\left[\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n = y_n)} = \pi_0^{(n)},$$

$$P(E_n=-1|Y_n=y_n) = \frac{\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_-^{(n-1)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\left[\lambda\mu_i^{(n-1)}-\Delta\right]\right)^2}{2\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n=y_n)} = \pi_-^{(n)},\ and$$

$$P(E_n=+1|Y_n=y_n) = \frac{\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_+^{(n-1)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\left[\lambda\mu_i^{(n-1)}+\Delta\right]\right)^2}{2\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n=y_n)} = \pi_+^{(n)}.$$

For proof of the theorem, see technical appendix -A.

Note that the formulas in this theorem are all recursive. In particular, $P(E_n \mid Y_n = y_n)$ is derived with the use of the prior at stage $n-1$ (that involves information about y_{n-1}) which is, on itself, an updated posterior coming from stage $n-2$ and so on. Therefore, historical data information is maintained and updated as each new observation y_n accrues. Other points that remain to be exposed from the theorem are the subscripts and the superscripts of the iterative weights, means and variances. The subscript i denotes each component of the mixture at stage $n-1$; $i \in \{0, 1, \ldots, 3^{n-1}-1\}$. As one moves from $n-1$ to n, each i triplicates, giving rise to a mixture of $3 \times 3^{n-1} = 3^n$ components. To keep track of the $flat-rise-decay$ cases and facilitate their recognition, the components are renumbered at time n using the module 3 operator, i.e.:

$3i+0$ refers to the flat state (corresponding to mod(3)=0)

$3i+1$ refers to the decay state (corresponding to mod(3)=1)

$3i+2$ refers to the rise state (corresponding to mod(3)=2).

Thus, the modulo 3 operator allows us to easily discriminate which of the components refers to what of the three possible states of the epidemic. The problem of monitoring the time-indexed series comes down to estimating and monitoring two of the three $\pi_\bullet^{(n)}$'s.

2.1 Drawing Decision Regarding the State of the Epidemic

In General: At each stage of the sequential algorithm, three actions may be taken. The actions $a_f^{(n)}$, $a_r^{(n)}$ and $a_d^{(n)}$ are respectively associated with the *flat, rise,* and *decay* states of the epidemic at stage n. Some of these actions may be: reinforcing current measures of preparedness, taking stronger eradicative measures, or allocating resources to other means. Depending on whether $\pi_\bullet^{(n)}$ (i.e. any state posterior probability at stage n) has actually increased, decreased, or stabilized, there would be some costs incurred. A loss matrix such

as in Table 1 can be built and decision rules for hypotheses testing can be worked into the sequential algorithm. Note that a Bayes rule always exists since the action space is a compact set.

	$a_d^{(n)}$	$a_f^{(n)}$	$a_r^{(n)}$
$E_n = -1$	0	$c_{-1,f}$	$c_{-1,r}$
$E_n = 0$	$c_{0,d}$	0	$c_{0,r}$
$E_n = +1$	$c_{1,d}$	$c_{1,f}$	0

Table 1: Loss matrix for a more general decision problem. In the matrix, $c_{\bullet,\bullet}$ represents the costs associated with the true state of E_n (denoted by the first subscript) and the action taken (second subscript).

Expert's opinion may be used to provide the appropriate values for the loss matrix. Once the matrix is set, one can derive the 'optimal' decision, according to some criterion, like Bayes risk, Minimax rule or other similar optimality criteria (see for example De Groot, 1970 and Berger, 1980).

Special case: In the following special case, we are interested in detecting the rise state of the epidemic (which in public health is usually of highest importance) while controlling the false alarm rate. This corresponds to the sequential hypothesis testing:

$$\left\{ \begin{array}{ll} H_0^{(n)}: & E_n = +1 \\ H_1^{(n)}: & E_n \neq +1 \end{array} \right\}.$$

Denote $a_0^{(n)}$ and $a_1^{(n)}$ the decisions to accept $H_0^{(n)}$ and $H_1^{(n)}$ respectively. Also, denote c_{MA} the cost associated with the 'Missed Alarm' (failing to recognize the *rise*, or Type I error); and c_{FA} the one associated with 'False Alarm' (falsely claiming a *rise*, or Type II error). The loss matrix becomes a 2×2 matrix, and according to the generalized 0–1 loss function takes the form of Table 2.

	$a_0^{(n)}$	$a_1^{(n)}$
$E_n = +1$	0	c_{MA}
$E_n \neq +1$	c_{FA}	0

Table 2: Loss matrix generated by the 0–1 generalized loss function principle for the special decision problem.

It is then easy to show (Casela and Berger, 1990) that the Bayes decision rule (i.e. the decision which minimizes Bayes risk) is: 'Reject $H_0^{(n)}$ if and only if'

$$P(E_n = +1|Y_n) < \frac{c_{FA}}{c_{MA} + c_{FA}}.$$

We should note here that other decision drawing methods such as Bayes factors (Jeffreys, 1948) are also plausible.

2.2 Thresholding the Posterior Probabilities

As noted earlier, the method is a sequentially updated recursive scheme where at stage n, the posterior probability of each state is used as prior for stage $n+1$. This technique invites a natural problem where one state with a history of consecutively low probabilities constantly contaminates its subsequent updates. This problem is known as causing sensitivity issues because a very low prior probability of an unlikely state would need several successive observations with high likelihood to provide an accurate signal. To avoid this problem, one usually puts a 'floor' on the value of a prior probability of any state; a technique that provides 'head start' to speed response to signal. The idea consists of defining a threshold floor value p^* such that $\pi_\bullet^{(n)} = \max\left\{p^*, \pi_\bullet^{(n)}\right\}$, and adjusting the remaining probabilities to attain collective exhaustiveness.

3 Attractiveness and Novelty of the Method

The attractiveness of the Bayesian method lies in its ease of adaptability to parameter instability such as the one seen in syndromic surveillance data settings. Bayesian approaches appear to be quite appealing in modeling these data because at least, the historical information (although usually imperfect) can be used to elicit an appropriate prior. Furthermore, in disease surveillance there is a need to draw inference in an online fashion even in a presence of a short data history. Our algorithm has the flexibility of a straightforward transition from fixed to a random jump. The sequential updating mechanism has allowed our approach to also perform well in the presence of a non informative prior for the state probabilities. Another appealing aspect is the possibility of deriving predictive distribution for future observable(s) to be used for error management or forecasting purposes. Specifically, following Geisser (1993) we easily show *via* some algebra (see technical appendix) that the predictive distribution of Y_n given the available data $Y_1, \ldots, Y_{n-1}$ is a mixture of 3^n Normal components:

$$
\begin{aligned}
f(Y_n|Y_{n-1}) &\sim \sum_{i\in\{0,-1,+1\}} f(Y_n|E_n=i)P(E_n=i|Y_{n-1}) \\
&\sim \sum_{i=0}^{3^{n-1}-1} \Bigg[w_i^{(n-1)}\pi_0^{(n-1)} N\left(\lambda\mu_i^{(n-1)},\ \ \sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right) \\
&+ \ w_i^{(n-1)}\pi_-^{(n-1)} N\left(\lambda\mu_i^{(n-1)}-\Delta,\ \ \tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right) \\
&+ \ w_i^{(n-1)}\pi_+^{(n-1)} N\left(\lambda\mu_i^{(n-1)}+\Delta,\ \ \tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right)\Bigg].
\end{aligned}
$$

This distribution opens the door to the predictive advantage of our mechanism.

We view our proposal as a self-reliant algorithm in that it combines three different aspects of control theory into a single process. It has the flexibility to simultaneously model, chart, and forecast. Modeling comes about by deriving the unconditional mixture distribution Y_n, sequential charting intervenes by taking advantage of the posterior state probabilities, and forecasting becomes apparent by making an adequate usage of the predictive distribution of Y_n.

3.1 Computational Considerations and Reproducibility

In an era of advanced computational technology, one may be tempted to implement our methodology in a *'brute-force'* manner. However, direct implementation of the mixtures involved in the theorem, and their sequential computation can lead to much severe computational load than necessary. The use of the three-state approach to modeling E_n induces a mixture of 3^n Normal components in the unconditional distribution of Y_n. Thus, we have a combinatorial explosion of the number of components used to describe the process. The reason for this explosion is that the model attempts to keep all possible trajectories from the beginning of the process till the most recent observation. Due to the autoregressive nature of the model, the posterior mixture parameters are bounded (see technical appendix A for these bounds). As n increases, the 3^n components have tiny weights and their parameters differ only slightly. Thus, we approximate the exact distribution with another mixture having far fewer components. We adopt the approximation algorithm of Tsiamyrtzis and Hawkins (2010), which is a hybrid of the procedure proposed in West (1993).

First select K, as the number of components for the approximating mixture. At stage n of the process, when the number of components $r = 3^n$ exceeds K for first time, activate the

approximation algorithm:

1. Sort the r components in increasing order of their weights.
2. Find the index i ($i = 2, 3, \ldots, r$) of the 'nearest neighbor' of component 1.
3. Replace these two distributions with a new Normal with updated parameters. Set $r = r - 1$.
4. Go to step 1 and repeat the procedure, until $r = K$.

Two issues that need to be clarified here are the definition of a 'nearest neighbor' (step 2) and the replacement of two Normal components by a single one (step 3).

To measure distribution proximity in finding the nearest neighbor (required at stage 2 of the algorithm), we use the Jeffreys (1948) divergence measure. Specifically, the Jeffreys divergence between f_1 and f_2 is provided by:

$$J(f_1, f_2) = \int [f_1(x) - f_2(x)] \log \left(\frac{f_1(x)}{f_2(x)} \right) dx.$$

In the case where $f_1 \sim N(\mu_1, \sigma_1^2)$ and $f_2 \sim N(\mu_2, \sigma_2^2)$ the criterion is given by:

$$J(f_1, f_2) = \frac{1}{2} \left(\frac{\sigma_1}{\sigma_2} - \frac{\sigma_2}{\sigma_1} \right)^2 + \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \frac{(\mu_1 - \mu_2)^2}{2}.$$

The lowest-weight mixture component is then merged with the component with lowest Jeffreys divergence, and the pair replaced with a single normal distribution. The mean and variance of the new component is estimated using the method of moments. Thus, if we have $f_1 \sim N(\mu_1, \sigma_1^2)$ and $f_2 \sim N(\mu_2, \sigma_2^2)$ with weights w_1 and w_2 respectively, the replacement will be given by $f_p \sim N(\mu_p, \sigma_p^2)$ with $w_p = w_1 + w_2$ and (see for example McLachlan and Basford 1988):

$$\mu_p = \left(\frac{w_1}{w_1 + w_2} \right) \mu_1 + \left(\frac{w_2}{w_1 + w_2} \right) \mu_2$$

$$\sigma_p^2 = \left(\frac{w_1}{w_1 + w_2} \right) \sigma_1^2 + \left(\frac{w_2}{w_1 + w_2} \right) \sigma_2^2 + \frac{w_1 w_2}{(w_1 + w_2)^2} (\mu_1 - \mu_2)^2,$$

which preserves the total weight, marginal mean and variance of the components. The choice of K would undoubtedly reflect on the required accuracy of the approximation, where the larger the K, the better the approximation. Based on West (1993) proposal, along with some simulation results, it turns out that usually a few hundred components suffice to approximate the exact posterior satisfactorily in most applications. We use $K = 1,000$ in our application; although $K = 500$ has a similar performance.

4 Simulation Study for Sensitivity Analysis

4.1 Performance Under Normal Regime

The performance under normal regime gives us an assessment of sensitivity analysis for real epidemics. Three different syndromes are used for the assessment; fever (FEV), respiratory illnesses 1 (RESP1) and respiratory illnesses 2 (RESP2). The clinical definition of these syndromes is attached in the technical appendix section B. These syndromes came from physicians visit records and emergency department discharges in a District of Columbia metropolitan area in the United State (U.S). The data was kindly supplied by Dr. Burkhom of the Applied Physics Laboratory at Johns Hopkins University. Observations are recorded daily over a span of 3 years (1994–1997) and represent an aggregate over multiple clinical settings. The hypothesis is that these variables may carry clinical information and symptoms that may indicate bioterrorism-associated diseases (see for example CDC-MMWR, vol 50, number 41, for a complete list and syndromic surveillance variables' rankings); or that an intentional release is likely to alter their daily readings. We use the first two years to elicit our nuisance parameters and check them against the third year. Table 3 gives us the estimated nuisance parameters.

To be noted too is the day of the week (DOW) effect usually depicted on these syndromic data. We reduce the DOW effect by using a seven-day rolling window filter. See Shmueli and Burkhom (2010); Lotze et al. (2008) for suggested methods for reducing a DOW effect. Other appropriate methods are available in spectral analysis; Bloomfield (1976).

We use the parameters in Table 3 to run our algorithm against the 1997 real-time epidemic curve to graphically assess its goodness of fit. In order for our approach to be used in biosurveillance, the algorithm should be able to model and follow the pathway of an epidemic curve. Since the underlying dual reason for using influenza-like-illnesses is that they correlate with influenza and that symptoms from the aftermath of biological activities are likely to reflect on them and alter their normal counts, we also found it appropriate to check our performance against the influenza pattern for a receiver operating characteristic (ROC) assessment. The gold standard for generating these ROC curves is the flu season which, for this specific location, ranges from November 1^{st} to March 31^{st} according to the U.S. CDC. This may differ from state to state and from one location to another.

Nuisance Parameters			
Parameters	FEV	RESP1	RESP2
λ	0.94	0.95	0.95
ζ	60.0	200	150
σ_0^2	16.0	220	80.0
σ^2	16.0	220	80.0
Δ	$4\sqrt{16}$	$3\sqrt{220}$	$3\sqrt{80}$
τ^2	2.00	16.0	9.00
$\pi_+^{(0)}$	0.20	0.20	0.20
$\pi_-^{(0)}$	0.20	0.20	0.20
p^*	0.10	0.10	0.10
K	1000	1000	1000

Table 3: Table of nuisance parameters. λ is the autoregressive discount factor; ζ, σ_0^2 are the initial prior mean and variance; Δ and τ^2 are the mean and variance of the random (in magnitude) jump distribution; $\pi_+^{(0)}$ ($\pi_-^{(0)}$) are the initial prior probability referring to rise (decay) states (for the flat state we have $\pi_0^{(0)} = 1 - \pi_+^{(0)} - \pi_-^{(0)}$), σ^2 is the model variance and p^* is the floor probability. K is the number of components used to approximate the mixtures.

Results:

Figures 1–3 show our model's approximation to the state and the stage of an epidemic curve. The model was run on the 1997 data set for FEV, RESP1, and RESP2. On left-hand panel of each graph, we display the unconditional mean Y_n as stated in *Theorem 1'* approximation of the epidemic stage (grey), and the actual data values (black). The graphs on the left-hand panel give graphical evidence that the unconditional mean Y_n provides a good fit to the observed readings. The right-hand panel displays two of the three posterior state probabilities. Observe that the posterior probabilities change direction relative to changes in the epidemic curve or with respect to changes in Y_n. This suggests their reliability for sensitivity analysis, provided they have an ease of predictability. This predictability is assessed in a ROC analysis. Figure 4. displays the ROC curves. The curves were built using $\pi_0^{(n)}$ as ROC variable (left-hand panel) and on the right-hand panel using the posterior $\pi_+^{(n)}/\pi_0^{(n)}$ as a ROC criterion. The ROC suggests that RESP1 unequivocally correlates better with influenza season and is closely followed by FEV. It also is intuitively clear from both panels that RESP2 may not be the ideal prognostic factor. The low performance of RESP2 may be attributed to the syndrome' assessment; although tuning parameters may also have a say. Turning to the ICD9 codes defining the syndrome, RESP2 is defined as other categories of respiratory illnesses that cannot be clearly identified as RESP1; either known or unknown. This catego-

rization becomes subjective by definition and invites more misclassification, irregularity and more structural break than the clearly defined FEV and RESP1.

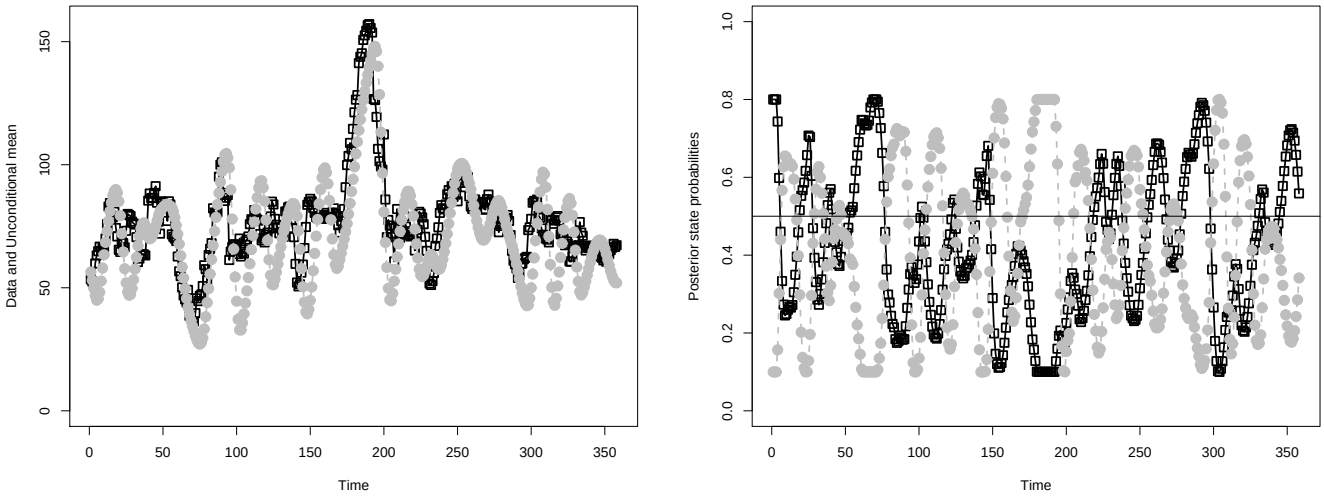

Figure 1: Year 1997; time origin July 1^{st}. Left panel: FEV data in black and squared dots; unconditional mean in grey and circles. Right panel: Posterior state probabilities; black and squared dots is $\pi_0^{(n)}$; and in grey $\pi_+^{(n)}$.

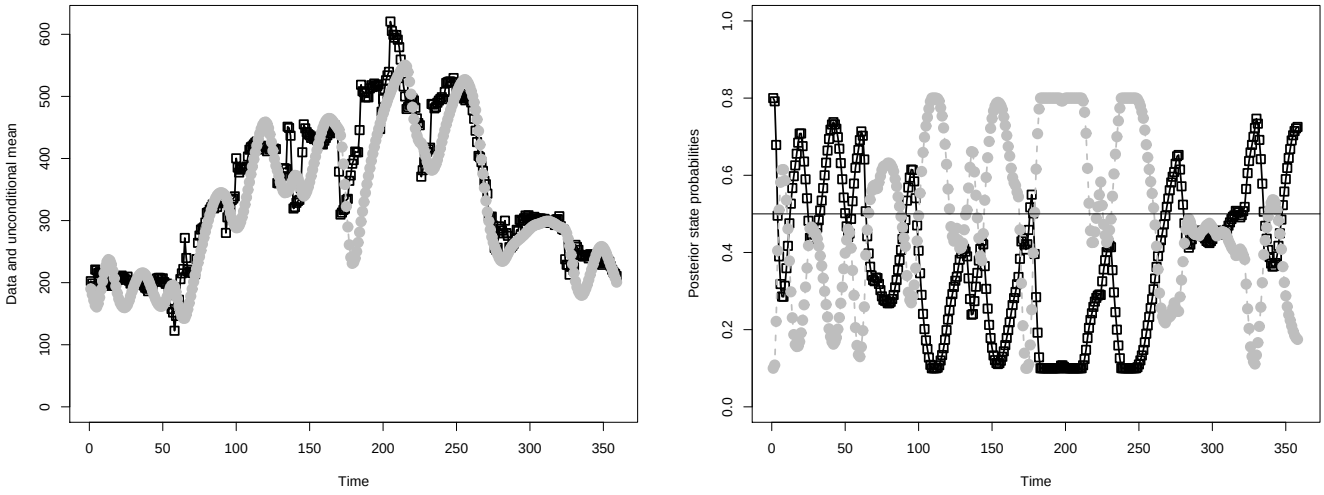

Figure 2: Year 1997; time origin July 1^{st}. Left panel: RESP1 data in black and squared dots; unconditional mean in grey and circles. Right panel: Posterior state probabilities; black and squared dots is $\pi_0^{(n)}$; and in grey $\pi_+^{(n)}$.

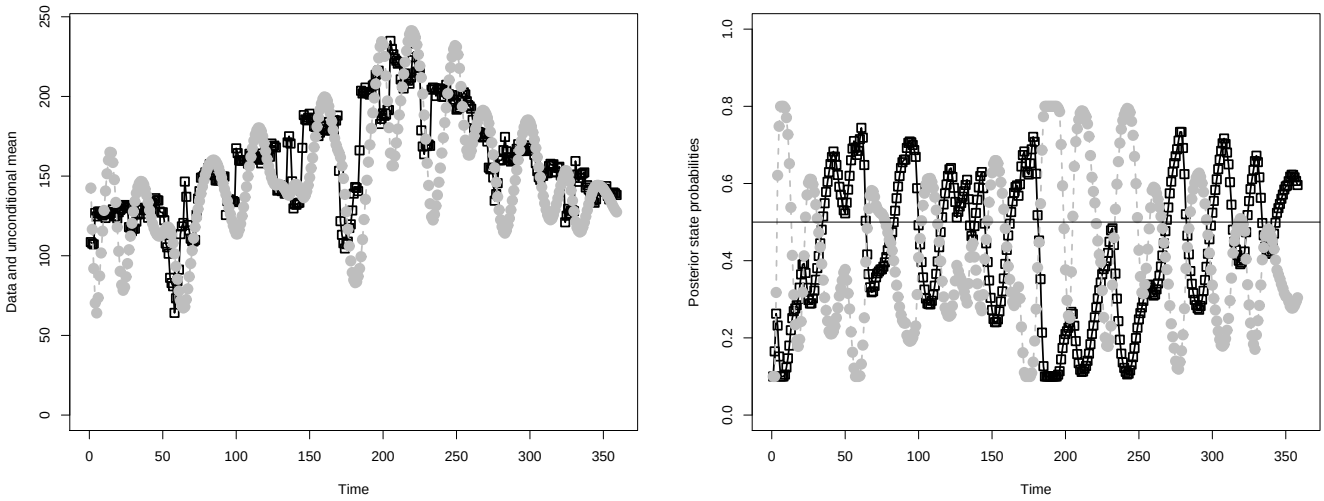

Figure 3: Year 1997; time origin July 1^{st}. Left panel: RESP2 data in black and squared dots; unconditional mean in grey and circles. Right panel: Posterior state probabilities; black and squared dots is $\pi_0^{(n)}$; and in grey $\pi_+^{(n)}$.

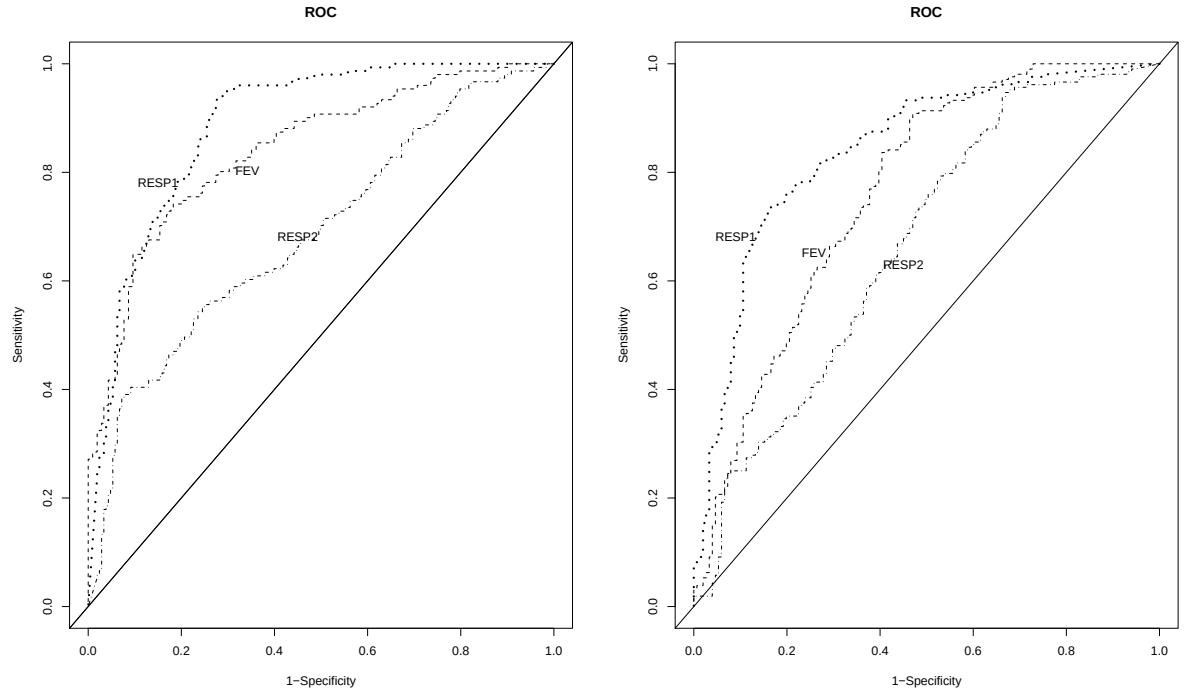

Figure 4: ROC curves for Fever (FEV) and Respiratory illnesses. Variable used for left-hand panel is $\pi_0^{(n)}$. Variable used for right-hand ROC panel is $\pi_+^{(n)}/\pi_0^{(n)}$. RESP1 dominates. In all, RESP2 is not particularly effective.

The ROC analysis is particularly useful in finding a threshold value $\hbar(\pi_{\bullet}^{(n)})$ that yields the 'best acceptable' set of sensitivity and specificity. For example, with $\hbar(\pi_0^{(n)}) = 0.42$, RESP1 yields a sensitivity of 92.71% and a specificity of 72.59%. With the same threshold FEV yields a sensitivity of 80.79% and a specificity of 73.55%. The value of $\hbar(\pi_0^{(n)})$ can be used as threshold limit for a natural outbreak. Should one define a cost or utility function that combines these two probability values, the 'best acceptable' threshold $\hbar(\pi_0^{(n)})$ may be chosen to minimize this utility function. (See for example Metz, 1978; Pepe, 2003). To chart and identify additional cases that could alter normal counts (e.g. in case of intentional release), it is intuitive to keep a copy of the charting parameters $(\pi_{\bullet}^{(n)})$ under normal regime and assess performance with respect to this baseline. In the absence of expert's opinion to conduct a full utility/cost analysis, we conduct a simulation study under normal regime to calibrate a false alarm rate for charting. We run 1000 replications in which the observed FEV (or RESP1) series is perturbed by normal fluctuation $N(0, \sigma^2)$. We assess false alarm relative to the original true series through numerical optimization. To be explicit, if $\pi_+^{(n)}/\pi_0^{(n)}$ is the ratio of the two posteriors under the true series, and $\pi_+^{(n)}/\pi_0^{(n)} \mid_s$ is the corresponding value in repeated iterations, we find a and b such that $P(a < \pi_+^{(n)}/\pi_0^{(n)} - \pi_+^{(n)}/\pi_0^{(n)} \mid_s < b) = 1 - \alpha$. With $a = -0.225$ and $b = 0.311$, the false alarm rate is $\alpha = 0.002$. This false alarm rate is carried into the following simulated outbreaks study for signal detection.

4.2 Performance Under Different Outbreak Scenarios

Since the problem that spearheaded this research is outbreaks signal detection problem, we simulated outbreaks in different ways that a pathogen could alter syndrome counts (see Mandl et al., 2004 and Figure 8 in appendix B). The simulated outbreak extra cases are superposed on the 1997 HDS and run through the developed model for signal detection. We simulated outbreaks of one and two weeks. We spike the data stream with a number of individual non-overlapping outbreaks. We space the outbreaks by 10, 20, and 30 days throughout the year to ensure that seasonality in the data does not impede signal detection. We clean the system before the next outbreak so that response to one outbreak does not carry over to the next. The simulated data represents multiple ways in which pathogens could spread through a community. Among the pathways considered, a flat outbreak corresponds to point-source infections such as Bacillus Anthracis. Linear, exponential and sigmoid outbreaks may relate to infectious diseases that are highly infectious, such as smallpox. Our working assumption is that these outbreaks will reflect upon FEV and RESP1 (we discard RESP2 because of its marginal performance in a ROC analysis). The number of simulated additional cases

depends on the error profile of the agent under study (twice the standard deviation). Two error profiles are considered; one for the flu season (winter) and one outside the flu season (summer). The reasoning behind this is that visit rates vary more unpredictably in winter than in summer. To illustrate, if the mean daily visit is 80 and the pathway gives a constant error profile with standard deviation of 25, simulated outbreaks ranging from 0 to 50 extra visits per day are considered. These visits are randomly sampled from 0 to 50 at a step of 5, and readjusted to fit the distributional pathway under study. In case the error profile may vary between summer and winter, the number of additional cases may follow the season accordingly, with more extra cases in winter. For example FEV has an error profile with standard deviation 30 in winter and 22 otherwise.

4.2.1 Simulation Setting

We simulated one thousand (1,000) outbreaks for each of the following combinations: duration (7 days, 14 days), pathway (flat, linear, exponential, sigmoid), and yearly hit frequency (once, twice). In case of more than one outbreak, the spacings used are 10, 20 or 30 days. We kept a copy of the posterior ratio under normal regime and compared the simulated performances to the normal for signal detection.

Here, we describe the basic idea behind generating a single outbreak. The idea carries over to multiple non-overlapping outbreaks.

1. Choose a random $n_1 \in \{1, \ldots, 365 - d\}$ for the beginning of an outbreak with duration d

2. Prior to n_1 there is a normal course of an epidemic curve. Thus, the simulated data before n_1 is the original data $1, \ldots, n_1 - 1$ plus a normal fluctuation $N(0, \sigma^2)$

3. From time n_1, generate an outbreak with duration d and pathway q. Add these outbreaks to the original data stream plus a normal fluctuation

4. After time $(n_1 + d)$, the normal course continues; so, add normal fluctuation to the original data set

5. Run the simulated data through the system for signal detection.

Figure 5 is an illustration of extra cases altering the normal FEV epidemic curve. For the sake of clarity of the figure, we ignore the random fluctuations. This provides a pictorial representation of how a single outbreak of duration d may look. In case of multiple epidemics,

one may use spacing after step 4, then cycle back to step 1 and choose a restricted $n_{1,2} >> n_1$. It would also be wise to confine n_1 to a time interval in case of multiple outbreaks.

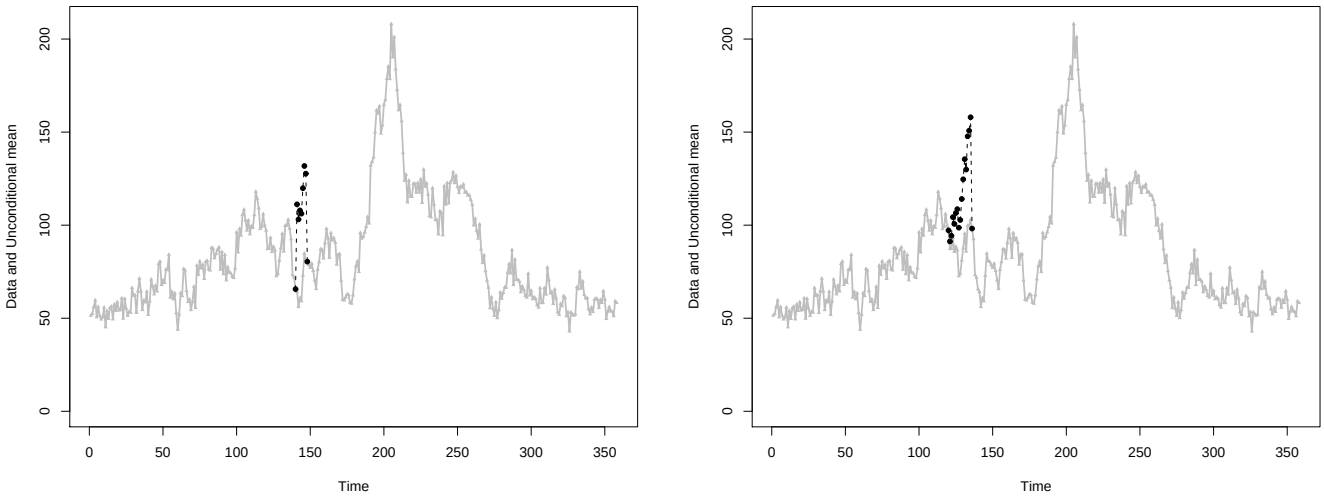

Figure 5: Left-hand panel: Epidemic curve (grey) spiked with a seven-day flat outbreak at day 148 (black). Right-hand panel: Epidemic curve (grey) spiked with a fifteen-day linear outbreak starting at day 122 (black). Time origin is July 1^{st}.

Assuming an outbreak was introduced at time n_1 and that our system flags contamination at time n_2, we define the following measuring yardsticks:

- No Signal (N.S.) cases where the system is degenerate and fails to sound an alarm after an outbreak has been introduced and during an outbreak
- Correct Signal (C.S.) cases where $n_2 = n_1$; to say that the system signals at the exact time that an outbreak has been introduced
- Missed Signal (M.S.) cases where $n_2 > n_1$; to say that the system misses the exact time but signals with delay
- False Signal (F.S.) cases where $n_2 < n_1$; to say that the system signals earlier than the event has actually occurred (this has been set to have probability 0.002).

These partitions provide a tool for performance assessment under unusual activity. In addition, a pictorial representation of signal and the missed signal compartments will help display the delay to signal.

4.2.2 Results of Sensitivity Analysis

Figures 6 – 7 provide a summary of the simulation results for single outbreaks. Among the series that did not result in a F.S. prior to spiking the data stream with simulated outbreaks, we evaluate the proportion of signal as well as the distribution of signals. Note that the flat pathway and the linear result in the best detection of signal (see table 4 & 5). On the tables, we display the cumulative signal detection as a numerical representation of the delay to signal. Note that the seven-day flat picks out nearly all its signals two days after being introduced (94.0%). The linear picks out nearly all its signals 3 days into spiking the data stream (95.2%). The seven-day exponential and sigmoid have a slower reaction time. This is to be expected because their shape starts out with low number of outbreaks then becomes gradually increasing (see appendix B for a display of possible shapes). Some series failed to signal during the outbreak period. We view those series as degenerate. Their proportions for the seven-day outbreak follow: flat (1.2%) linear (1.6%), sigmoid (3.2%), and exponential (2.8%). The proportion of degenerate series fades when the outbreak has a long duration (see Table 5). The non-overlapping outbreaks simulation scenario yields similar results. This is not a surprise because the system has been cleaned moving from the first to the second outbreak.

% of Signal	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Flat	79.6	92.4	94.0	96.0	97.2	98.4	98.8
Linear	29.2	71.2	91.2	95.2	97.2	98.0	98.4
Sigmoid	15.6	35.6	73.6	91.6	95.6	96.6	96.8
Exponential	07.2	16.4	38.8	71.2	87.6	95.6	97.2

Table 4: Table of cumulative signal for 7-day outbreaks. System set at a false alarm rate of 0.002

% of Signal	Day 2	Day 4	Day 6	Day 8	Day 10	Day 12	Day 14
Flat	92.8	96.8	97.6	98.0	99.2	99.2	100
Linear	37.2	87.2	96.4	98.0	98.4	99.6	100
Sigmoid	08.0	14.4	23.6	47.2	88.8	98.8	100
Exponential	12.8	20.8	27.6	32.4	62.8	94.8	99.2

Table 5: Table of cumulative signal for 14-day outbreaks. System set at a false alarm rate of 0.002

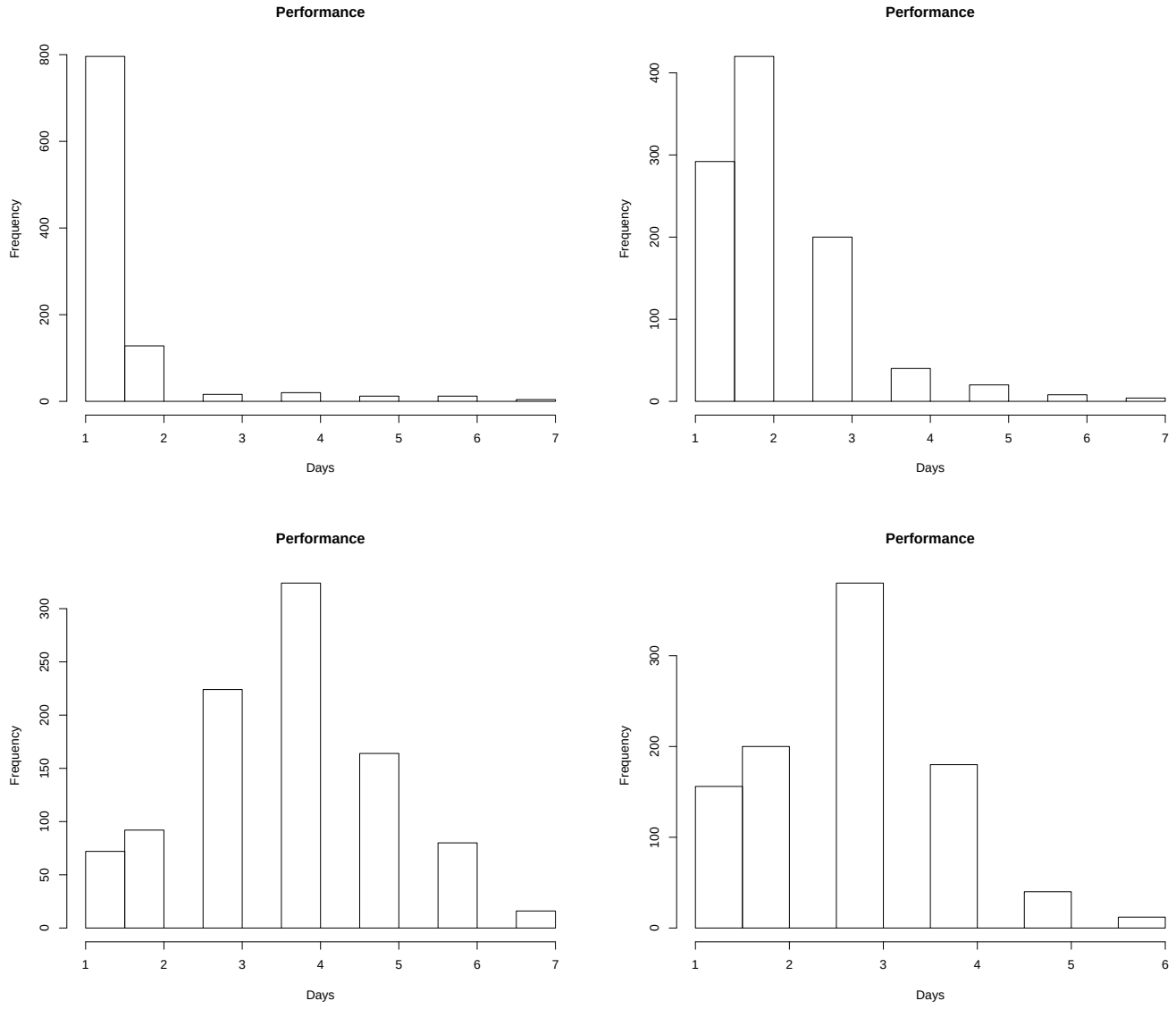

Figure 6: Performance on 7-day outbreaks. Top row: Flat and linear Linear. Bottom row: Exponential and Sigmoid.

5 Simulation Comparative Study

Since part of our goal is to build a control algorithm with the potential of identifying clusters of disease cases before traditional methods, we use the decision interval Cusum on the first difference readings following the 7-day rolling window filtration. The reason for using the Cusum on the first difference spurred from the fact that first differences have a tendency to remove bias from series and leave them with noise. After all, the readings have to be detrended/demodulated for the Cusum to reach its optimality. Next, the first differences are

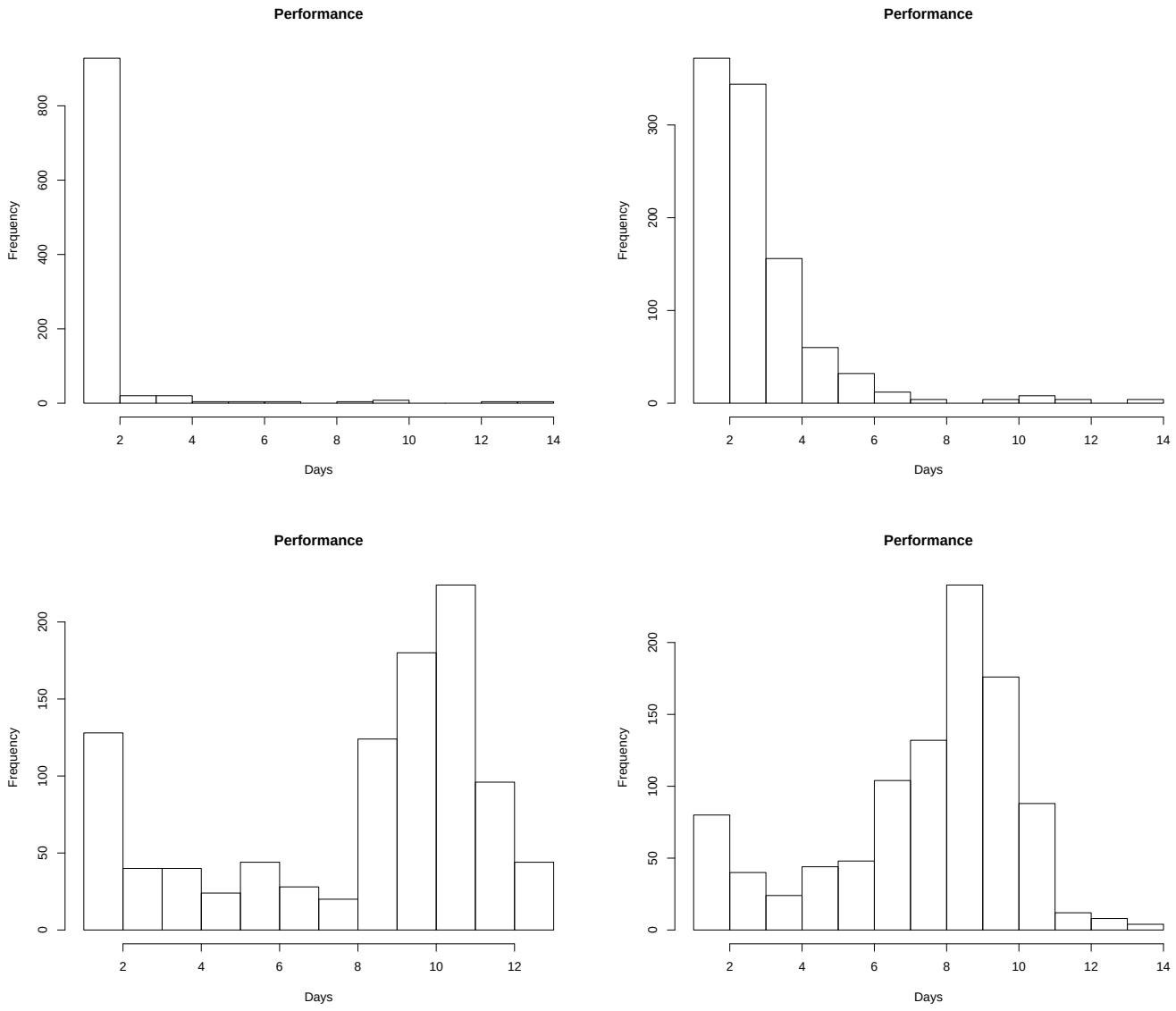

Figure 7: Performance on 14-day outbreaks: Top row: Flat and linear Linear. Bottom row: Exponential and Sigmoid.

transformed to achieve normality. Our approach builds 2 Cusums for shift in the mean; one upward and one downward. We run these Cusums until one crosses its decision interval, so that the out of control run length is the minimum of the two. We compare this first-signal-based Cusum to the performance of the Bayesian approach. To be fair, the comparison is plausible on one ground only: the flat outbreak pathway. The reason is that optimality of a Cusum holds from one step change to another step change; not from step to gradual changes. It would have been a clear disadvantage for the Cusum if the comparison were carried for outbreaks with gradual pathway. Speaking of the Cusum design, in order to obtain a joint

in-control ARL of 500, we must tune each individual Cusum at a nominal ARL of 1000. The in-control ARL was matched by the performance of the Bayesian method under normal operation resulting in a false alarm rate of 0.002 (see Montgomery 2009, Hawkins and Olwel 1998). Since we used 2-year observations to elicit the Bayesian model parameters, the same number of observations is used to calibrate our Cusums. The filtered and demodulated HDS is N(0.02, 4.71); which needed another standardization. The reference value and the decision intervals for the Cusums depend on the shift we anticipate to detect. In our case, with reference value of 1 for an upward shift, -1 for a downward shift, and a decision interval of 2.6666, both Cusums are tuned to detect 2 standard deviation shifts from a N(0,1) error profile and attain an average run length of 1000 each. We run 1000 of contaminated series through both the Bayesian model and through the Cusum; and we recorded the proportion of correct signals and that of degenerated series. Tables 6 and 7 compare the two methods and the signal detection efficiency of the Bayesian over the Cusum–which is in the order of 1.5 to 1. Note that in the Cusum case, more than 30% of the series degenerate. The intent of this comparative study is by no means to downgrade cumulative sums; but rather to bring to light some of its abuses. The point is that a process loses its signal detection capability when readings are extensively filtered; see Zhang (1997).

% of Signal	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Bayesian (Ba)	79.6	92.4	94.0	96.0	97.2	98.4	98.8
Cusum (Cu)	53.6	62.0	62.8	63.6	64.4	65.2	65.2
Efficiency (Ba/Cu)	1.48	1.49	1.49	1.50	1.50	1.50	1.51

Table 6: 7-day flat outbreaks. Comparing the Bayesian algorithm to the Cusum performance

% of Signal	Day 2	Day 4	Day 6	Day 8	Day 10	Day 12	Day 14
Bayesian (Ba)	92.8	96.8	97.6	98.0	99.2	99.2	100
Cusum (Cu)	60.0	62.4	63.6	66.0	66.8	68.4	68.8
Efficiency (Ba/Cu)	1.54	1.55	1.53	1.48	1.48	1.45	1.45

Table 7: 14-day flat outbreaks. Comparing the Bayesian algorithm to the Cusum performance

6 Discussion & Conclusion

Traditional SPC tools for monitoring data pertaining to biosurveillance are currently short of capturing all the important structural changes depicted on these data. These tools lose their

performance due to the background of variability seen on biosurveillance data. We have developed a three-state sequential and recursive algorithm that accounts for some of the issues encountered in biosurveillance. The algorithm is bayesian and accounts for non-stationarity, irregularity, and seasonality through sequentially updated prior distribution parameters. Our approach has the added advantage to adjust to random jumps, both in magnitude and in direction, and to capture an epidemic curve serial structural details. Compared to Cusum on simulated flat outbreaks, its efficiency is 1.5 to 1.

Other approaches based on Hidden Markov Models (HMM) and Markov Switching Models (MSM) are also used in conjunction with epidemiological types of data. Within this framework, an unobserved epidemic state space is modeled using homogeneous Markov chain of order one *(1)* with stationary transition probabilities (see for example Le Strat et al., 1999, for HMM and Lu et al., 2010, for MSM; see also Cappé, et al., 2007 and Früehwirth-Schnatter, 2006 for statistical developments in HMM and MSM models). These methods are usually computationally intensive with a questionable performance in start-up settings when only a short history is available at the beginning of a monitoring scheme. Also, multivariate methods may be worth exploring when many syndromic variables are under study. This issue needs more space and mathematical development than can be devoted to it here. Multivariate case charts are not assumption-free and are more demanding in their calibration needs than univariates'. A reasonable multivariate development would consist of using some syndromic variables in a 'strength borrowing' fashion to monitor the remaining incidence parameters.

The current proposal will augment early reports of natural epidemics or intentional operations to sentinels. Its novelty comes from monitoring the states posterior probabilities and from its self-reliance. We view our proposal as a self-reliant method, with the potential to combine three different aspects of control theory into a single process: modeling, charting, and forecasting. We also recognize that our proposal is not *'one size fits all'*. Biosurveillance data may differ from one location to another in the same country. As one switches from location to location or between syndromic variables, the general philosophy behind our proposal holds; but parameters may change. Thus, the need to re-tune some parameters and reset some initial values.

7 Technical Appendix–A

The theorem is proved by induction. For $n = 1$ it is easy to show that the theorem holds. Specifically, the unconditional distribution of Y_1 is a mixture of 3 Normal distributions:

$$f(Y_n) = \sum_{i=0}^{2} w_i^{(1)} N\left(\mu_i^{(1)}, \left(\sigma_i^{(1)}\right)^2\right)$$

where for the weights, means and variances, we have:

$$\begin{array}{rclcrclcrcl} w_0^{(1)} & = & \pi_0^{(0)} & & \mu_0^{(1)} & = & \lambda\zeta & & \left(\sigma_0^{(1)}\right)^2 & = & \sigma^2 + \lambda^2\sigma_0^2 \\ w_1^{(1)} & = & \pi_-^{(0)} & & \mu_1^{(1)} & = & \lambda\zeta - \Delta & & \left(\sigma_1^{(1)}\right)^2 & = & \tau^2 + \sigma^2 + \lambda^2\sigma_0^2 \\ w_2^{(1)} & = & \pi_+^{(0)} & & \mu_2^{(1)} & = & \lambda\zeta + \Delta & & \left(\sigma_2^{(1)}\right)^2 & = & \tau^2 + \sigma^2 + \lambda^2\sigma_0^2 \ . \end{array}$$

For the posterior probabilities of the events E_i we have:

$$P(E_1 = 0|Y_1 = y_1) = \frac{\sum_{i=0}^{2} \frac{\pi_0^{(0)}}{\sqrt{2\pi\left[\sigma^2+\lambda^2\sigma_0^2\right]}} exp\left\{-\frac{(y_1-\lambda\zeta)^2}{2\left[\sigma^2+\lambda^2\sigma_0^2\right]}\right\}}{f(Y_1 = y_1)} = \pi_0^{(1)},$$

$$P(E_1 = -1|Y_1 = y_1) = \frac{\sum_{i=0}^{2} \frac{\pi_-^{(0)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\sigma_0^2\right]}} exp\left\{-\frac{(y_1-[\lambda\zeta-\Delta])^2}{2\left[\tau^2+\sigma^2+\lambda^2\sigma_0^2\right]}\right\}}{f(Y_1 = y_1)} = \pi_-^{(1)}, and$$

$$P(E_1 = +1|Y_1 = y_1) = \frac{\sum_{i=0}^{2} \frac{\pi_+^{(0)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\sigma_0^2\right]}} exp\left\{-\frac{(y_1-[\lambda\zeta+\Delta])^2}{2\left[\tau^2+\sigma^2+\lambda^2\sigma_0^2\right]}\right\}}{f(Y_1 = y_1)} = \pi_+^{(1)} \ .$$

Next, assume that the theorem holds for $n-1$; i.e. the unconditional distribution of Y_{n-1} given by

$$f(Y_{n-1}) = \sum_{i=0}^{3^{n-1}-1} w_i^{(n-1)} N\left(\mu_i^{(n-1)}, \left(\sigma_i^{(n-1)}\right)^2\right),$$

where for the weights, means, variances, and the posterior probabilities of the events E_i, the recursive relationships are valid. Then, we can show that it is valid for n. Indeed,

$$\begin{array}{rcl} Y_n|Y_{n-1}, \delta, E_n & \sim & N\left(\lambda Y_{n-1} + E_n\delta, \ \ \sigma^2\right) \\ Y_{n-1} & \sim & \sum_{i=0}^{3^{n-1}-1} w_i^{(n-1)} N\left(\mu_i^{(n-1)}, \ \ \left(\sigma_i^{(n-1)}\right)^2\right) \\ \delta & \sim & N\left(\Delta, \ \ \tau^2\right). \end{array}$$

From the above we obtain the distribution of $Y_n|E_n$:

$$\begin{aligned} Y_n|E_n &\sim \sum_{i=0}^{3^{n-1}-1} w_i^{(n-1)} N\left(\lambda\mu_i^{(n-1)} + E_n\Delta,\ \ E_n^2\tau^2 + \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2\right) \\ E_n &= \left\{\begin{array}{ccc} 0 & \text{with prob.} & \pi_0^{(n-1)} \\ -1 & \text{with prob.} & \pi_-^{(n-1)} \\ +1 & \text{with prob.} & \pi_+^{(n-1)} \end{array}\right\}. \end{aligned}$$

Then for the unconditional distribution of Y_n we have:

$$\begin{aligned} Y_n &\sim \sum_{i=0}^{3^{n-1}-1} \left[w_i^{(n-1)}\pi_0^{(n-1)} N\left(\lambda\mu_i^{(n-1)},\ \ \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2\right)\right. \\ &+ w_i^{(n-1)}\pi_-^{(n-1)} N\left(\lambda\mu_i^{(n-1)} - \Delta,\ \ \tau^2 + \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2\right) \\ &+ \left. w_i^{(n-1)}\pi_+^{(n-1)} N\left(\lambda\mu_i^{(n-1)} + \Delta,\ \ \tau^2 + \sigma^2 + \lambda^2\left(\sigma_i^{(n-1)}\right)^2\right)\right] \\ &\equiv \sum_{i=0}^{3^n-1} w_i^{(n)} N\left(\mu_i^{(n)},\ \ \left(\sigma_i^{(n)}\right)^2\right); \end{aligned}$$

which satisfies the recursive relationships for the weights, means and variances. Furthermore, for the posterior distribution of E_n we get:

$$P(E_n = 0|Y_n = y_n) = \frac{\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_0^{(n-1)}}{\sqrt{2\pi\left[\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\lambda\mu_i^{(n-1)}\right)^2}{2\left[\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n = y_n)} = \pi_0^{(n)},$$

$$P(E_n = -1|Y_n = y_n) = \frac{\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_-^{(n-1)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\left[\lambda\mu_i^{(n-1)}-\Delta\right]\right)^2}{2\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n = y_n)} = \pi_-^{(n)},$$

$$P(E_n = +1|Y_n = y_n) = \frac{\sum_{i=0}^{3^{n-1}-1} \frac{w_i^{(n-1)}\pi_+^{(n-1)}}{\sqrt{2\pi\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}} exp\left\{-\frac{\left(y_n-\left[\lambda\mu_i^{(n-1)}+\Delta\right]\right)^2}{2\left[\tau^2+\sigma^2+\lambda^2\left(\sigma_i^{(n-1)}\right)^2\right]}\right\}}{f(Y_n = y_n)} = \pi_+^{(n)};$$

which are consistent with the formulas provided in the theorem.

Bounds on the parameters of $f(Y_n)$:

Here, we establish the proof that the means, variances and weights are bounded.

Means

The lowest (highest) possible mean corresponds to the case where we have a negative (positive) jump at each of the n stages of the process. Using induction one has:

$$\begin{aligned}
\mu^{(n)}_{min} &= \lambda^n\zeta - \lambda^{n-1}\Delta - \lambda^{n-2}\Delta - \ldots - \lambda\Delta - \Delta \\
&= \lambda^n\zeta - \left(\frac{1-\lambda^n}{1-\lambda}\right)\Delta, \\
\mu^{(n)}_{max} &= \lambda^n\zeta + \lambda^{n-1}\Delta + \lambda^{n-2}\Delta + \ldots + \lambda\Delta + \Delta \\
&= \lambda^n\zeta + \left(\frac{1-\lambda^n}{1-\lambda}\right)\Delta.
\end{aligned}$$

Thus at each stage n of the process all 3^n means are bounded in

$$\left[\mu^{(n)}_{min},\ \mu^{(n)}_{min}\right] = \left[\lambda^n\zeta - \left(\frac{1-\lambda^n}{1-\lambda}\right)\Delta,\quad \lambda^n\zeta + \left(\frac{1-\lambda^n}{1-\lambda}\right)\Delta\right],$$

with range converging as follow:

$$2\left(\frac{1-\lambda^n}{1-\lambda}\right)\Delta \xrightarrow{n\uparrow} \frac{2\Delta}{1-\lambda}.$$

As n grows, these means are within the range of at most $2\Delta/(1-\lambda)$, leading to the conclusion that they become less and less different from each other as n increases.

Variances

The two most extreme scenarios for the variances are:

$\left(\sigma^{(n)}_{min}\right)^2$: corresponding to a no-jump state for all n stages of the process

$\left(\sigma^{(n)}_{max}\right)^2$: corresponding to all-jump states (positive or negative) for all n stages of the process. Based on the recursion we can easily show:

$$\begin{aligned}
\left(\sigma^{(n)}_{min}\right)^2 &= \sigma^2 + \lambda^2\sigma^2 + \lambda^4\sigma^2 + \ldots + \lambda^{2n-2}\sigma^2 + \lambda^{2n}\sigma_0^2 \\
&= \lambda^{2n}\sigma_0^2 + \left(\frac{1-\lambda^{2n}}{1-\lambda^2}\right)\sigma^2, \\
\left(\sigma^{(n)}_{max}\right)^2 &= \left(\sigma^2+\tau^2\right) + \lambda^2\left(\sigma^2+\tau^2\right) + \lambda^4\left(\sigma^2+\tau^2\right) + \ldots + \lambda^{2n-2}\left(\sigma^2+\tau^2\right) + \lambda^{2n}\sigma_0^2 \\
&= \lambda^{2n}\sigma_0^2 + \left(\frac{1-\lambda^{2n}}{1-\lambda^2}\right)\left(\sigma^2+\tau^2\right).
\end{aligned}$$

At each stage n of the process all 3^n variance terms are bounded in

$$\left[\left(\sigma^{(n)}_{min}\right)^2,\ \left(\sigma^{(n)}_{max}\right)^2\right] = \left[\lambda^{2n}\sigma_0^2 + \left(\frac{1-\lambda^{2n}}{1-\lambda^2}\right)\sigma^2,\quad \lambda^{2n}\sigma_0^2 + \left(\frac{1-\lambda^{2n}}{1-\lambda^2}\right)\left(\sigma^2+\tau^2\right)\right],$$

with range converging as follow:

$$\left(\frac{1-\lambda^{2n}}{1-\lambda^2}\right)\tau^2 \xrightarrow{n\uparrow} \frac{\tau^2}{1-\lambda^2}.$$

As n increases the variance terms become similar.

Weights

The weights involve a non-linear relation between the posterior probabilities and the data readings. An exact mathematical bound on these entities in not a reasonable task. However, given that the weights fall in [0,1] and add up to 1, as n increases all 3^n weights become smaller. Based on the recursive relationship at each stage n of the process the weights of the previous stage $n-1$ are multiplied with the respective posterior probabilities (i.e. a number in [0,1]) making them smaller.

Technical Appendix–B

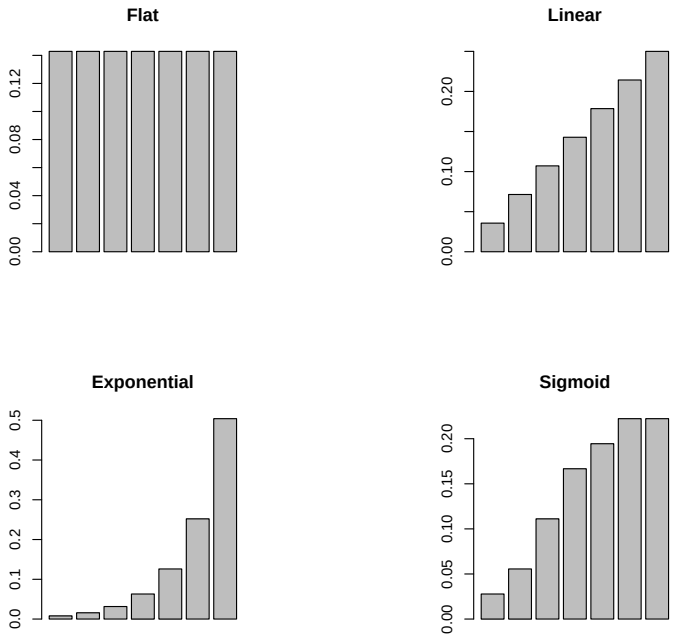

Figure 8: Various shapes of outbreaks considered. These are canonical frequency shapes for display; random displacements have not been added to them yet; the span of each outbreak is 7 days.

FEV: Includes febrile illnesses of unspecified origin, unknown viral illnesses accompanied by fever, fever and septicemia not specified.
RESP: Includes any acute infection of the upper and/or lower respiratory tract (from the oropharynx to the lungs, includes otitis media), diagnosis of acute respiratory tract infection such as pneumonia due to parainfluenza virus, acute non-specific diagnosis of respiratory tract infection such as sinusitis, pharyngitis, laryngitis, cough, stridor, shortness of breath, and throat pain. Exclude chronic bronchitis, allergic conditions and asthma. These respiratory diagnoses are believed to be linked to anthrax (inhalational) , tularemia and plague (pneumonic).

RESP1: Reflects general symptoms of respiratory illnesses and also includes associated bioterrorism diseases of highest concern or diseases highly approximating them.

RESP2: Consists of symptoms that might normally be placed in respiratory illness group but daily volume could detract from the signal generated by RESP1.

8 Reference

Arnon S.S., Schechter R., Inglesby T.V., 2001. Tulinum toxin as a biological weapon: medical and public health management. Journal of the American Medical Association. 285, 1059–70.

Berger, J.O., 1980. Statistical Decision Theory; Foundation, Concept and Methods, Springer Verlag, New York.

Bloomfiled, P., 1976. Fourier Analysis for Time Series: An Introduction, Wiley, New York.

Buehler J.W., Berkelman R.L., Hartley D.M., Peters C.J., 2003. Syndromic surveillance and bioterrorism-related epidemics. Emerging Infectious Diseases. 9, 1197-1204.

Cappé, O., Moulines, E., Rydén, T., 2007. Inference in Hidden Markov Models, Springer Series in Statistics, Springer, New York.

Casella, G., Berger, R. L., 1990. Statistical Inference, Wadsworth & Brooks-Cole, Pacific Grove, California.

Center for Disease Control and Prevention, MMWR 2001. Recognition of Illness Associated with the Intentional Release of a Biologic Agent. MMWR, 10.19.2001, 50(41), 893- 897.

De Groot, M.H., 1970. Optimal Statistical Decision, McGraw-Hill, New York.

Dennis D.T., Inglesby T.V., Henderson D.A., 2001. Tularemia as a biological weapon: medical and public health management. Journal of the American Medical Association. 285, 2763–73.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models, Springer Series in Statistics, Springer, New York.

Geisser, S., 1993. Predictive Inference: An Introduction, Chapman & Hall, London.

Green M.S., Kaufman Z., 2002. Surveillance for early detection and monitoring of infection disease outbreaks associated with bioterrorism. Israel Medical Association Journal. 4(7), 503-6.

Hawkins, D.M., Olwell, D.H., 1998. Cumulative Sum Charts and Charting for Quality Improvement, Springer Verlag, New York.

Hawkins, D.M., Qiu, P., Kang, C.W., 2003. The Change point Model for Statistical Process Control. Journal of Quality Technology. 35, 355-365.

Hawkins, D.M., Zamba, K.D., 2005a. A Change point Model for Statistical Process Control with Shift in mean or Variance. Technometrics. 47(2), 164-173.

Hawkins, D.M., Zamba, K.D., 2005b. Change point Model for a Shift in Variance. Journal of Quality Technology. 37(1), 21-31.

Henderson D.A., Inglesby T.V., Bartlett J.G., 1999. Smallpox as a biological weapon: medical and public health management. Journal of the American Medical Association. 281, 2127–37.

Hutwagner L.C., Maloney E.K., Bean N.H., Slutsker L., Martin S.M., 1997. Using Laboratory-based Surveillance data for Prevention: An algorithm to detect Salmonella Outbreaks. Emerging Infectious disease. 3, 395-400.

Jeffreys, H., 1948. Theory of Probability, Second Edition, University Press, Oxford.

Inglesby T.V., Henderson D.A., Bartlett J.G., 1999. Anthrax as a biological weapon: medical and public health management. Journal of the American Medical Association. 281, 1735.

Inglesby T.V., Dennis D.T., Henderson D.A., 2000. Plague as a biological weapon: medical and public health management. Journal of the American Medical Association. 283, 2281–2290.

Le Strat, Y., Carrat, F., 1999. Monitoring epidemiologic surveillance data using hidden Markov models. Statistics in Medicine. 18, 34633478.

Lotze, T., Murphy, S., Shmueli, G., 2008. Implementation and Comparison of Preprocessing Methods for Biosurveillance Data. Advances in Disease Surveillance. 6(1), 1-20.

Lu, H.-M., Zeng, D., Chen, H., 2010. Prospective infectious disease outbreak detection using Markov switching models. IEEE Transactions on Knowledge & Data Engineering. 22, 565577.

Mandl, K.D., Reis, B., Cassa, C., 2004. Measuring Outbreak–Detection Performance by Using Controlled Feature Set Simulations. Morbidity and Mortality Weekly Report (MMWR). 53, 130–136.

McLachlan, G.J., Basford, K.E., 1988. Mixture Models Inference and Applications to Clustering, Marcel Dekker, New York.

Metz, C.E., 1978. Basic Principle of ROC Analysis. Seminars in Nuclear Medicine. 8, 283–298.

Montgomery, D.C., 2009. Introduction to Statistical Quality Control, John Wiley and Son, New York.

Pavlin J.A., 2003. Investigation of Disease Outbreaks Detected by Syndromic Surveillance Systems. Journal of Urban Health. 80, 107-114.

Pepe, M.S., 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction, University Press, Oxford.

Shmueli, G., Burkom, H.S., 2010. Statistical Challenges Facing Early Outbreak Detection in Biosurveillance. Technometrics (Special Issue Anomaly Detection). 52(1), 39-51.

Stern, L., Lightfoot D., 1999. Automated outbreak detection: A Quantitative Retrospective Analysis. Epidemiology of Infectious disease. 122, 103-110.

Stoto M.A., Schonlau M., Mariano L.T., 2004. Syndromic Surveillance: Is it Worth the Effort?. Chance. 17(1), 19-24.

Tsiamyrtzis, P., Hawkins, D.M., 2010. Bayesian Start up Phase Mean Monitoring of an Autocorrelated Process that is Subject to Random Sized Jumps. Technometrics. 52(4), 438-452.

West, M., 1993. Approximating posterior distributions by Mixtures. Journal of Royal Statistical Society, Series B. 55, 2, 409-422.

Winkel, P., Zhang, N.F., 2007. Statistical Development of Quality in Medicine, John Wiley & Sons Ltd, England.

Woodall, W.H., 2006. The Use of Control Charts in Health-Care and Public-Health Surveillance. Journal of Quality Technology. 38(2), 89–104.

Zamba, K.D., Tsiamyrtzis, P., Hawkins, D.M., 2008. A Sequential Bayesian Control Model for Influenza-Like Illnesses and Early Detection of Intentional Outbreaks. Quality Engineering. 20(4) 495-507.

Zhang, N.F., 1997. Detection capability of residual control chart for stationary process data. Journal of Applied Statistics. 24, 475-492.