

# Wavelet-based Monitoring for Modern Biosurveillance

Galit Shmueli

Dept of Decision & Information Technologies

Robert H. Smith School of Business

University of Maryland, College Park 20742

## **Abstract**

Current biosurveillance relies on classical statistical control charts for detecting disease outbreaks. However, these are not always suitable in this context. Assumptions of normality, independence, and stationarity are typically violated in syndromic data. Furthermore, outbreak signatures in such data are of unknown patterns, and therefore call for “general detectors”. We propose wavelet-based methods, which make less assumptions and are suitable for detecting abnormalities of unknown form. Wavelets have been widely used for data denoising and compression, but little work exists on using them for monitoring. We discuss monitoring-based issues and illustrate them using data on military clinic visits.

*Keywords:* Early detection; autocorrelation; disease outbreak; syndromic data; discrete wavelet transform.

# 1 Introduction to modern Biosurveillance

Biosurveillance is the practice of monitoring data for the purpose of detecting outbreaks of an epidemic. Traditional biosurveillance has focused on the collection and monitoring of medical and public health data that verify the existence of a disease outbreaks. Examples are laboratory reports and mortality rates. Although such data are the most direct indicators of a disease, they tend to be collected, delivered, and analyzed days, weeks, and even months after the outbreak. By the time this information reaches decision makers it is often too late to treat the infected population or to react in some other way. Modern biosurveillance has therefore adopted the notion of “syndromic data” in order to achieve early detection. Syndromic data include information such as over-the-counter and pharmacy medication sales, calls to nurse hotlines, school absence records, web-searches on medical websites, and chief complaints by individuals who visit hospital emergency rooms. All these do not directly measure an infection, but it is assumed that they contain an earlier, though weaker, signature of a disease outbreak. The various data sources fall along a continuum according to their “earliness”. Under the assumption that people tend to self-treat and self-medicate themselves before rushing to the hospital, we expect web searching and the purchasing of OTC medication to precede calls to nurse hotlines and ambulance dispatches, and then to be followed by emergency room visits. Still, this entire continuum is assumed to occur before actual diagnoses can be made (after hospitalization and/or lab tests).

Recent years have shown great improvement in the collection and transfer of syndromic data. Currently, many surveillance systems that are deployed across the country routinely collect data from multiple sources on a daily basis, and these data are transferred with very little delay to the biosurveillance systems (see Fienberg and Shmueli (2005) for a description of this process and examples from several surveillance systems). This means that the multiple syndromic data streams that are now collected are very different in nature than traditional data. We now have multiple time series that come from different sources, and therefore have different background behavior. Since these sources are believed to indirectly capture a disease outbreak, they tend to have a lower signal-to-noise ratio than traditional diagnostic data. This means that the noise includes sources of variation that are irrelevant to outbreak detection, such as the correlation of sales of cough medication with overall grocery sales. Almost all

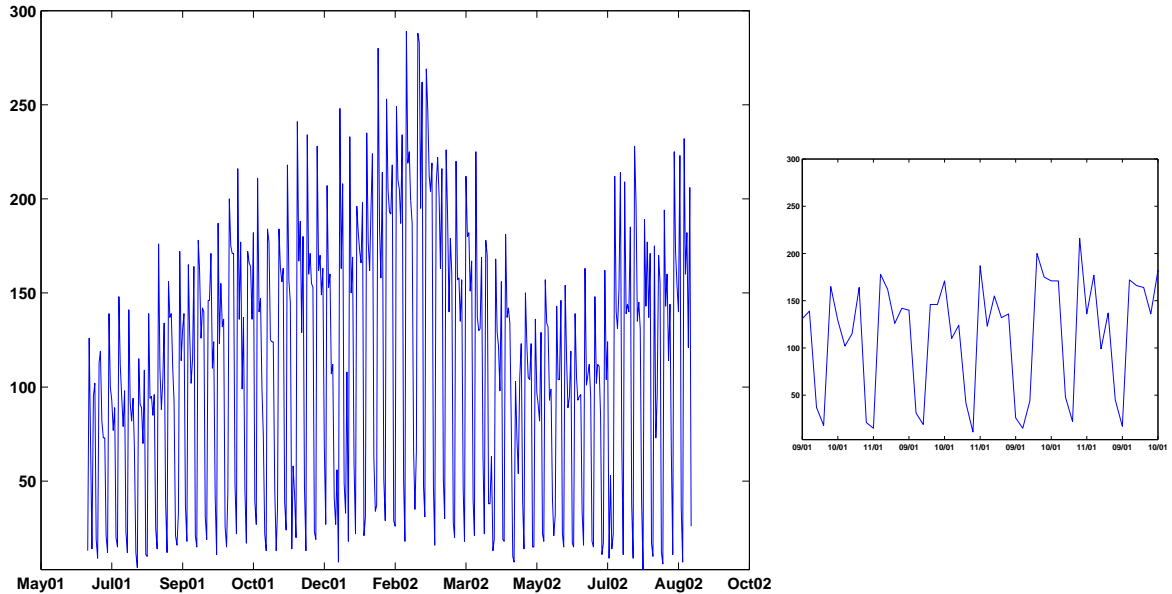


Figure 1: Daily number of respiratory chief complaints among visits to military outpatient clinics in Charleston between July 2001 and August 2002 (left) and zoomed-in view during Sept-2001 (right).

series exhibit some type of seasonality, which can differ across sources. The daily frequency of collection leads to non-negligible autocorrelation, and the nature of these non-specific datasets can include missing values (e.g. school absences are missing on holidays), coding errors, etc. Furthermore, the level of stationarity varies across series and across sources. These types of complications were demonstrated for over-the-counter medication sales (Goldenberg et al., 2002; Fienberg and Shmueli, 2005). Figure 1 illustrates some of these features. The left panel describes the daily number of visits to military outpatient clinics in Charleston that resulted in respiratory complaints. The noisiness and changing mean is visible. During this period epidemiologists believe that an outbreak took place between Feb-19 and Mar-20 of 2002. We describe the data, outbreak, and further details in Section 3. Zooming in (right panel in Figure 1) reveals a clear day-of-week effect, with lower counts on weekends.

A further complication in biosurveillance is that the signature of a disease or bio-terror attack outbreak in syndromic and especially non-traditional data is of unknown pattern. We do not know how an Anthrax attack will manifest itself in ambulance dispatches or in sales of cough remedies. Also, these systems are currently being used for the dual purpose of detecting natural disease outbreaks and bio-terrorist attacks, which can be of very different nature. Due to this signature uncertainty it is

reasonable that methods that are tuned to particular anomaly patterns (such as classic control charts) will be on average less powerful than “general detectors”.

In the absence of syndromic data that contain actual bio-terrorist attacks, there is the additional ambiguity of how to evaluate the performance of different algorithms. Even for natural outbreaks, the assertion of when exactly an outbreak occurred is ambiguous. A recent biosurveillance program by the The Defense Advanced Research Projects Agency (DARPA) was aimed at evaluating different algorithms using a common set of syndromic data in multiple US cities. For determining natural outbreaks in the data a team of epidemiologists and medical specialists were assigned the task of identifying outbreaks. According to Siegrist et al. (2005) and as described in Siegrist and Pavlin (2004), the team used three methods to determine “golden-standards”: documented outbreaks identified by traditional surveillance, visual analysis of the data, and a simple statistical algorithm to identify anomalies in the data. This, of course, raises issues of how outbreaks and their dates are determined, the circularity of determining outbreaks by employing statistical surveillance, and the evaluation of actual detection rates, false alarm rates, and timeliness. This adds an additional layer of uncertainty that further distinguishes between biosurveillance and classic engineering process control. An alternative approach has been to seed syndromic data with an artificial outbreak. However, there is a major challenge in constructing realistic outbreak patterns because the signatures of such outbreaks in syndromic data are yet unknown.

Current surveillance systems rely mostly on traditional statistical monitoring methods such as statistical process control and regression-based methods (e.g., Burkom et al., 2004; Heffernan et al., 2004). The simplicity and familiarity of these methods to the public health community have led them to continue being implemented even in this new environment. However, it appears that more suitable monitoring methods should evolve, as they have in other fields where the data environment has change in this way. This has motivated our investigation of wavelet methods, which have been adopted in many fields, and in particular are suitable for the type of data and problem in biosurveillance.

## 2 Wavelet based monitoring

In this section we focus on how wavelets can be used for the purpose of monitoring, and what are important issues that arise. We start with a brief description of wavelets and the wavelet decomposition, focusing on the parts that are relevant for monitoring. There is a large literature on wavelets and their uses, and this is beyond the scope of this work.

Wavelets are a method for representing a time series in terms of coefficients that are associated with a particular time and a particular frequency (Percival and Walden, 2000). The wavelet decomposition is widely used in the signal processing world for denoising signals and recovering underlying structure. Unlike other popular types of decompositions such as the Fourier transform, the wavelet decomposition yields localized components. A Fourier transform decomposes a series into a set of sines and cosines, where the different frequencies are computed globally from the entire duration of the signal, thereby maintaining specificity only in frequency. In contrast, the wavelet decomposition offers a localized frequency analysis. It provides information not only on what frequency components are present in a signal, but also when, or where they are occurring (Abramovich et al., 2000). The wavelet filter is long in time when capturing low-frequency events, and short in time when capturing high-frequency events. This means that it can expose patterns of different magnitude and duration in a series, while maintaining their exact timing.

Wavelet decompositions have proven especially useful in applications where the series of interest is not stationary. This includes long-range dependent processes which include many naturally occurring phenomena such as river flow, atmospheric patterns, telecommunications, astronomy, and financial markets (Gencay et al., 2001, Chap 5). Also, the wavelet decomposition highlights inhomogeneity in the variance of a series. Furthermore, data from most practical processes are inherently multiscale due to events occurring with different localizations in time, space, and frequency (Aradhye et al., 2003). With today's technology, many industrial processes are measured at very high frequencies, yielding series that are highly correlated and sometimes non-stationary. Together with the advances in mechanized inspection, this allows much closer and cheaper inspection of processes, if the right statistical monitoring tools are applied. The appeal of Shewhart charts has been its simplicity, both in understanding and

implementing them, and its power to detect abnormal behavior when the underlying assumptions hold. On the other hand, the growing complexity of the measured processes has led to the introduction of various methods that can account for features such as autocorrelated measurements, seasonality, and non-normality. The major approach has been to model the series by accounting for these features, and monitoring the model residuals using ordinary control charts. One example is fitting ARIMA models to account for autocorrelation (Reis and Mandl, 2003). A more popular approach is using regression-type models that account for seasonality, e.g. the Serfling model (Serfling, 1963) that is widely used by the Centers for Disease Control and Prevention (CDC). A detailed modelling approach is practical when the number of series to be monitored is small, and when the model is expected to be stable over time. However, in many cases there are multiple series, each of differing nature that might change their behavior over time. In such cases fitting a customized model to each process and updating the model often is not practical. This is the case of biosurveillance.

An alternative approach is to use a more flexible set of monitoring tools that do not make as many restricting assumptions. Methods that fall into this category are batch-means control charts (Runger and Willemain, 1995), moving centerline EWMA (Mastrangelo and Montgomery, 1995), and dynamic PCA (Ku et al., 1995). These methods all attempt to account for autocorrelation in the data. The wavelet decomposition is also such an approach, but it makes even less assumptions: it can handle autocorrelation and non-stationarity.

An additional appeal of wavelet methods is that the nature of the abnormality need not be specified apriori. Unlike popular control charts such as Shewhart, CuSum, MA, and EWMA charts that operate on a single scale and are most efficient at detecting a certain type of abnormality (Box and Luceo (1997)), wavelets operate on multiple scales simultaneously. In fact, the multiscale SPC (MSSPC) wavelet-based method (Aradhye et al. (2003)) subsumes the Shewhart, MA, EWMA, and CuSum charts. Scales are inversely proportional to frequencies, and thus cruder scales are associated with higher frequencies.

## 2.1 The discrete wavelet transform (DWT)

The discrete wavelet transform is an orthonormal transform of a real-valued time-series  $\mathbf{X}$  of length  $n$ . Using the notation of Percival and Walden (2000), we write  $\mathbf{w} = W\mathbf{X}$ , where  $\mathbf{w}$  is an  $n \times 1$  vector of

DWT coefficients, and  $W$  is an  $n \times n$  orthonormal matrix defining the DWT. The time series can be written as

$$X_t = \sum_{j=1}^J D_{j,t} + A_{J,t} \quad (1)$$

where  $\mathbf{D}_j = W_j' w_j$  is the detail associated with changes in  $\mathbf{X}$  at scale  $j$ , and  $A_{J,t}$  is equal to the sample mean. This representation is called the multiresolution analysis (MRA) of  $\mathbf{X}$ .

The process of performing DWT to decompose a time series and the opposite reconstruction step are illustrated in Figure 2. The first scale is obtained by filtering the original time-series through a high-pass (wavelet) filter, which yields the first-scale detail coefficient vector  $cD_1$ , and through a low-pass (scaling) filter, which gives the first-scale approximation coefficient vector  $cA_1$ . In the next stage the same operation is performed on  $cA_1$  to obtain  $cA_2$  and  $cD_2$ , and so on. This process can continue to produce as many as  $J = \log_2(n)$  scales. In practice, however, a smaller number of scales is used. The standard DWT includes a step of down-sampling after each filtering, such that the filtered output is subsampled by two. This means that the number of coefficients reduces from scale to scale by a factor of 2. In fact, the vector of wavelet coefficients,  $\mathbf{w}$ , can be organized into  $J + 1$  vectors:  $\mathbf{w} = [cD_1, cD_2, \dots, cD_J, cA_J]$ , where  $cD_j$  and  $cA_j$  correspond to the detail and approximation coefficients at scale  $j$ , respectively. To reconstruct the series from its coefficients, the opposite operations are performed: starting from  $cD_J$  and  $cA_J$ , an upsampling step is performed (where zeros are introduced between each two coefficients), and “reconstruction filters” (or “mirror filters”) are applied in order to obtain  $cD_{J-1}$  and  $cA_{J-1}$ . This is repeated until the original series is finally reconstructed from  $cD_1$  and  $cA_1$ .

Notice that there is a distinction between the wavelet approximation and detail coefficients and the wavelet reconstructed coefficients, which are often called Approximations and Details. We denote the coefficients by  $cA$  and  $cD$  (corresponding to approximation and detail coefficients, respectively), and their reconstruction by  $A$  and  $D$ . This distinction is important, because some methods operate directly on the coefficients, whereas other methods use the reconstructed vectors. The reconstructed approximation  $A_j$  is obtained from  $cA_j$  by applying the same operation as the series reconstruction, with the exception that it is not combined with  $cD_j$ . An analogous process transforms  $cD_j$  to  $D_j$ .

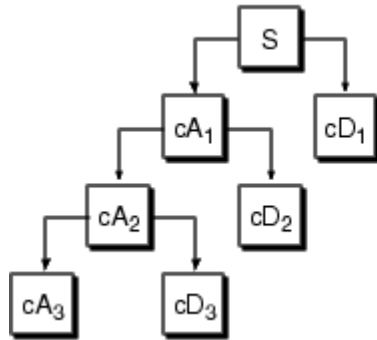


Figure 2: The process of DWT (from Matlab's Wavelet Toolbox Tutorial). The time series (S) is decomposed into 3 scales.

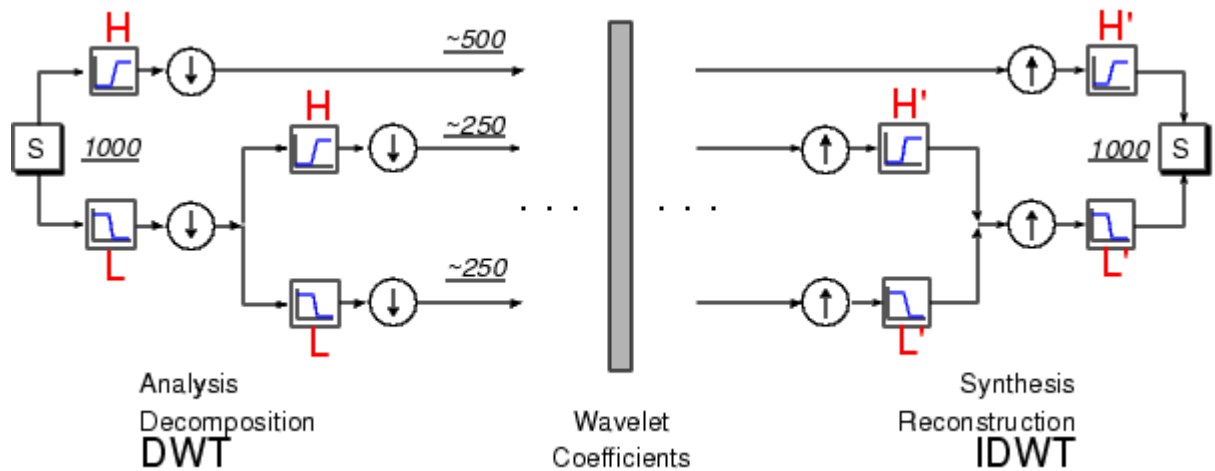


Figure 3: Decomposition (left) and reconstruction (right) of a signal to/from its coefficients. Decomposition is done by convolving the signal (S) of length  $n = 1000$  with a high-pass filter (H) and low-pass filter (L), and then downsampling. The next scale is similarly obtained by convolving the first level approximation with these filters, etc. Reconstruction is achieved by upsampling and then convolving the upsampled vectors with “mirror filters”. (from Matlab's Wavelet Toolbox Tutorial).

To illustrate the complete process, consider the Haar, which is the simplest wavelet. In general, the Haar uses two operations: averaging and taking differences. The low-pass filter takes averages ( $g = \sqrt{2}[1/2, 1/2]$ ), and the high-pass filter takes differences ( $h = \sqrt{2}[1/2, -1/2]$ ). Starting with a time series of length  $n$ , we obtain the first level approximation coefficients  $cA_1$  by taking averages of every pair of adjacent values in the time series (and multiplying by a factor of  $\sqrt{2}$ ). The first-level detail coefficients  $cD_1$  are proportional to the differences between these pairs of adjacent values. Downsampling means that we drop every other coefficient, and therefore  $cA_1$  and  $cD_1$  are each of length  $n/2$ . At the next level,  $cA_2$  coefficients are obtained by averaging adjacent pairs of  $cA_1$  coefficients. This is equivalent to averaging groups of four values in the original time series.  $cD_2$  coefficients are obtained from  $cA_1$  by taking the difference between adjacent pairs of  $cA_1$  coefficients. This is equivalent to taking the difference between pairwise-averages of the original series. Once again, a downsampling step removes every other coefficient, and therefore  $cA_2$  and  $cD_2$  are each of length  $n/4$ . These operations are repeated to obtain the next scales. In summary, the detail coefficient at scale  $j$ , at time  $t$  reflects the difference between averages of  $2^j - 1$  values before and after time  $t$  (the time series is considered to be at scale  $j = 0$ ).

The Haar reconstruction filters are given by  $g^* = -g$ , and  $h^* = h$ . These are used for reconstructing the original series and the approximation and details from the coefficients.

Figures 4-5 shows the result of a DWT with 5 levels for the (deseasonalized) daily number of visits to military outpatient clinics dataset (deseasonalization is described further in Section 3). Figure 4 displays the Multiresolution Analysis which includes the reconstructed coefficients ( $D_1 - D_5, A_5$ ). Figure 5 shows the set of 5 detail coefficients ( $cD_1, \dots, cD_5$ ) in the bottom panel, and the 5th level approximation ( $A_5$ ) overlaid on top of the original series.

The next step is to utilize the DWT for monitoring. There have been a few different versions of how the DWT is used for the purpose of forecasting or monitoring. In all the underlying idea is to decompose the signal using DWT and then operating on the individual detail and approximation levels. Some methods operate on the coefficients (e.g., Aradhye et al., 2003), while others use the reconstructed approximation and details (Renaud et al., 2005). There are a few other variation, such as Zhang et al.

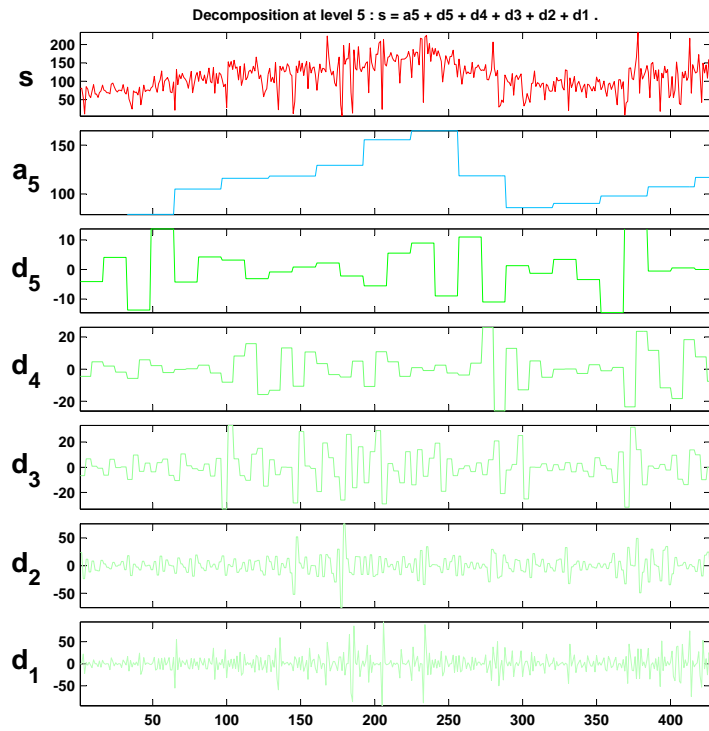


Figure 4: Decomposing the series of daily visits to military outpatient clinics into 5 levels using the Haar wavelet. The top panel displays the time series. Below it is the multiresolution analysis that includes the 5th approximation and the 5 levels of detail.

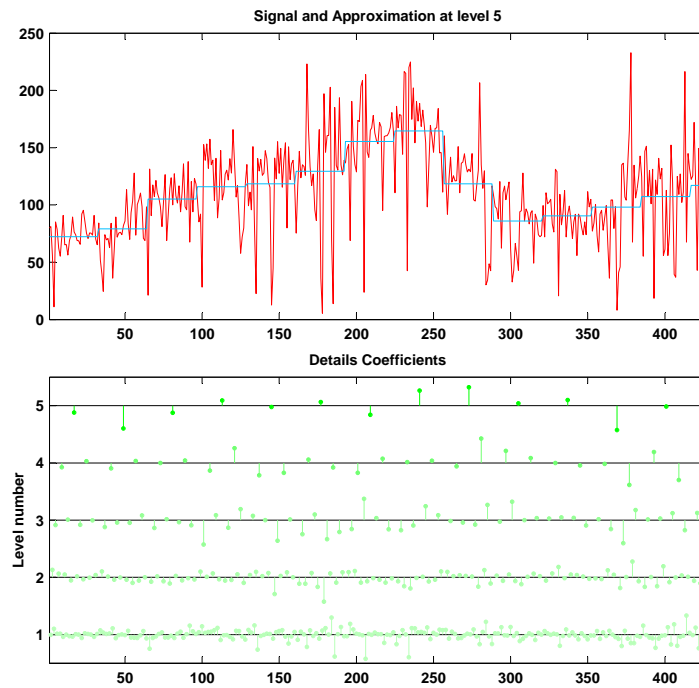


Figure 5: Decomposing the (deseasonalized) series of daily visits to military outpatient clinics into 5 levels using the Haar wavelet. The top panel displays the time series overlaid with the 5th approximation (the blocky curve). The bottom panel shows the 5 levels of detail coefficients.

(2003) who use the crudest approximation to de-trend the data in order to remove extremely low data points due to missing data. Our goal is to describe a general approach for **monitoring** using a wavelet decomposition. We therefore distinguish between retrospective and prospective monitoring, where in the former past data are examined to find anomalies retrospectively and in the latter the monitoring is done in real-time for new incoming data, in an attempt to detect anomalies as soon as possible. Although this distinction has not been pointed out clearly, it turns out that the choice of method, its interpretation, and performance are dependent on the goal.

## 2.2 Retrospective monitoring

Using wavelets for retrospective monitoring is the easier task of the two. The goal is to find anomalous patterns at any time points within a given time series. This can be done by constructing thresholds, or “control limits” that distinguish between natural variation and anomalous behavior. Wang (1995) suggested the use of wavelets for detecting jumps and cusps in a time series. He shows that if these patterns are sufficiently large, they will manifest themselves as extreme coefficients in the high detail levels. Wang uses Donoho’s universal threshold to determine what an extreme coefficient is. This threshold is given by  $\sigma\sqrt{(2)\log(n)/n}$ , where  $\sigma$  is estimated by the median absolute deviation of the coefficients at the finest level, divided by 0.6745 (Wang, 1995).

Another approach that is directly aimed at statistical monitoring was suggested by Bakshi (1998). The method, called Multiscale Statistical Process Control (MSSPC), comprises of three steps:

1. Decomposing the time series using DWT
2. Using a Shewhart (or other) control chart to threshold coefficients at each of the details scales and at the coarsest approximation scale, such that coefficients within the control limits are zeroed out. Control limits are computed from scale-specific coefficients.
3. Reconstructing the time series from the thresholded coefficients, and monitoring this series using a Shewhart (or other) control chart. The control limits at time  $t$  are based on estimated variances from scales where coefficients exceeded their limits at time  $t$ .

The method relies on the decorrelation property of DWT, such that coefficients within and across scales are approximately uncorrelated. It is computationally cheap and easy to interpret. Using a set of real data from a chemical engineering process, Aradhya et al. (2003) show that this method exhibits better average performance compared to traditional control charts in detecting a range of anomalous behavior when autocorrelation is present or void. Its additional advantage is that it is not tailored to detecting a certain type of anomaly (e.g., single spike, step function, exponential drift), and therefore on average performs better than pattern-specific methods.

MSSPC is appealing because its approach is along the line of classical SPC. However, there are two points that need careful attention: the establishment of the control limits, and the triggering of an alarm. We describe these next.

### 2.2.1 Two-phase monitoring

We put MSSPC into the context of two-phase monitoring: Phase I consists of establishing control limits based on a period with no anomalies. For this purpose, there should be a period in the data that is known to be devoid of anomalies. Phase II uses these control limits to detect abnormalities in the remainder of the data. Therefore a DWT is performed twice: once for the purpose of estimating standard deviations and computing control limits (using only the in-control period), and once for the entire series, as described in steps 2-3.

### 2.2.2 Accounting for Multiple Testing

MSSPC triggers an alarm only if the reconstructed series exceeds its control limits. The reconstruction step is essentially a denoised version of the original signal with hard thresholding. An alternative, which does not require the reconstruction step and directly accounts for the multiple tests at the different levels, is described below. When each of the levels is monitored at every time point, a multiple testing issue arises: For an  $m$ -level DWT there are  $m + 1$  tests being performed at every time point in step 2. This results in an inflated false alarm rate. When these tests are independent (as is the case because of the decorrelation property) and each has a false alarm rate  $\alpha$ , the combined false alarm rate at any time point will be  $1 - (1 - \alpha)^{m+1}$ . We suggest correcting for this multiple testing by integrating

the False Discovery Rate (FDR) correction (Benjamini and Hochberg, 1995; Storey, 2002) into step 2 above. Unlike Bonferroni-type corrections that control for the probability of any false alarm, the FDR correction is used to control the average proportion of false alarms among all detections. It is therefore more powerful in detecting real outbreaks. Adding an FDR correction means in practice that instead of setting fixed control limits, the p-values are calculated at each scale, and the thresholding then depends on the collection of p-values.

Finally, using DWT for retrospective monitoring differs from ordinary control charts in that it compares a certain time point not only to its past, but also to its future. Depending on the wavelet function the type and strength of relationship with neighboring points (e.g., symmetric vs. non-symmetric wavelets and the width of the wavelet filter).

### 2.3 Prospective monitoring

The use of DWT for prospective monitoring is more challenging. A few wavelet-based prospective monitoring algorithms were suggested and applied in biosurveillance (e.g., Goldenberg et al., 2002; Zhang et al., 2003). However, there has not been a rigorous discussion of the statistical challenges associated with the different formulation. We therefore list some main issues and suggest some possible solutions.

1. **Time lag:** Although DWT can approximately decorrelate highly correlated series (even long memory with slow decaying autocorrelation function), the downsampling introduces a time lag for calculating coefficients, which becomes more and more acute at the coarse scales. One solution is to use a “redundant” or “stationary” DWT (termed SWT), where the downsampling step is omitted. The price is that we lose the decorrelation ability. The detail and approximation coefficients in the SWT are each a vector of length  $n$ , but they are no longer uncorrelated within scales. The good news is that the scale-level correlation is easier to handle than the correlation in the original time-series, and several authors have shown that this dependence can be captured by computationally efficient models. For example, Goldenberg et al. (2002) use a redundant DWT and scale-specific AR models to forecast one-step-ahead details and approximation. These

forecasts are then combined to create a one-step-ahead forecast of the time series. Aussem and Murtagh (1997) use a similar model, with neural networks to create the one-step-ahead scale-level forecasts. From a computational point of view, the SWT computational complexity is  $O(n \log_2 n)$  like that of the FFT algorithm, compared to  $O(n)$  for DWT. So it is still very reasonable.

2. **Dependence on starting point:** Another challenge is the dependence of the decomposition on what we take as the starting point of the series (Percival and Walden, 2000). This problem does not exist in the SWT, which further suggests the adequacy of the redundant DWT for prospective monitoring.
3. **Boundaries:** DWT suffers from “boundary problems”. This means that the beginning and end of the series need to be extrapolated somehow in order to compute the coefficients. The amount of required extrapolation depends on the filter length and on the scale. The Haar is the only wavelet that does not impose boundary problems for computing DWT coefficients. Popular extrapolation methods are zero-padding, mirror- and periodic-borders. In the context of prospective monitoring of a time-series, the boundary problem is acute, because it weakens the most recent information that we have. To reduce the impact of this problem, it is better to use wavelets with very short filters (e.g., the Haar which has length 2), and to keep the depth of the decomposition to a minimum. Notice that an SWT using the Haar does suffer from boundary problems, although minimally ( $2^j$  values at scale  $j$  are affected by the boundary). Renaud et al. (2005) and Xiao et al. (2005) suggest a “backward implemented” SWT that uses a non-symmetric Haar filter for the purpose of forecasting a series using SWT with the Haar. This implementation only uses past values of the time series to compute coefficients. Most wavelet functions, excluding the Haar, use future values to compute the coefficients at time  $n$ . Examining this modification, we find that it is equivalent to using an ordinary SWT except that coefficients at scale  $j$  are shifted in time by  $2^j$  points. For instance, the level 1 approximation and detail coefficients will be aligned with the second time point, rather than the first time point. This means that at scale  $j$  the first  $2^j - 1$  coefficients are missing, so that we need a phase I that is longer than  $2^J$ . The great advantage of this modification is that new data-points do not affect past coefficients. This means that we can

use this very efficiently in a roll-forward algorithm that applies SWT with every incoming point.

These challenges all suggest that a reasonable solution is to use a Haar-based redundant DWT in its “backward” adaptation as a basis for prospective monitoring. This choice also has the advantage that the approximation and details are much smoother than the downsampled Haar DWT, and therefore more appealing for monitoring time series that do not have block-type structure. Furthermore, the redundant-DWT does not require the time series to be of special length, as might be required in the ordinary DWT. Finally, the Haar is the simplest wavelet function, and is therefore desirable from a computational and parsimonious perspective. Figure 6 illustrates the result of applying a “backward” redundant Haar wavelet decomposition to the military clinic visits data. It can be seen that the series of coefficients starts at different time-points for the different scales. However, for prospective monitoring this is not important, because we assume that the original time series is sufficiently long and that our interest focuses on the right-most part of the series.

Two-stage monitoring should also be implemented in this case: first, a period that is believed to be devoid of any disease outbreak of interest is used for estimating scale-specific parameters. Then, these estimates are used for determining control limits for alerting of anomalous behavior.

The redundancy in wavelet coefficients now leads to an inflated false alarm rate. Aradhya et al. (2003) approached this by determining empirically the adjustment to the control limits in MSSPC when using the redundant DWT. The disadvantage of using the redundant DWT is that the high autocorrelation in the coefficients increases the false alarm rate in highly non-stationary series (Aradhya et al., 2003). An FDR correction can still be used to handle the correlation across scales. In the following we suggest an approach to handling the autocorrelation within scales directly.

### **2.3.1 Handling scale-level autocorrelation**

As can be seen in Figure 6, the coefficients within each scale are autocorrelated, with the dependence structure being different for the detail coefficients compared with the approximation coefficients series. In addition to autocorrelation, sometimes after deseasonalizing there is still a correlation at lag 7. This type of periodicity is common in many syndromic datasets. The underlying assumption behind the detail coefficients is that in the absence of anomalous behavior they should be zero. From our experience with

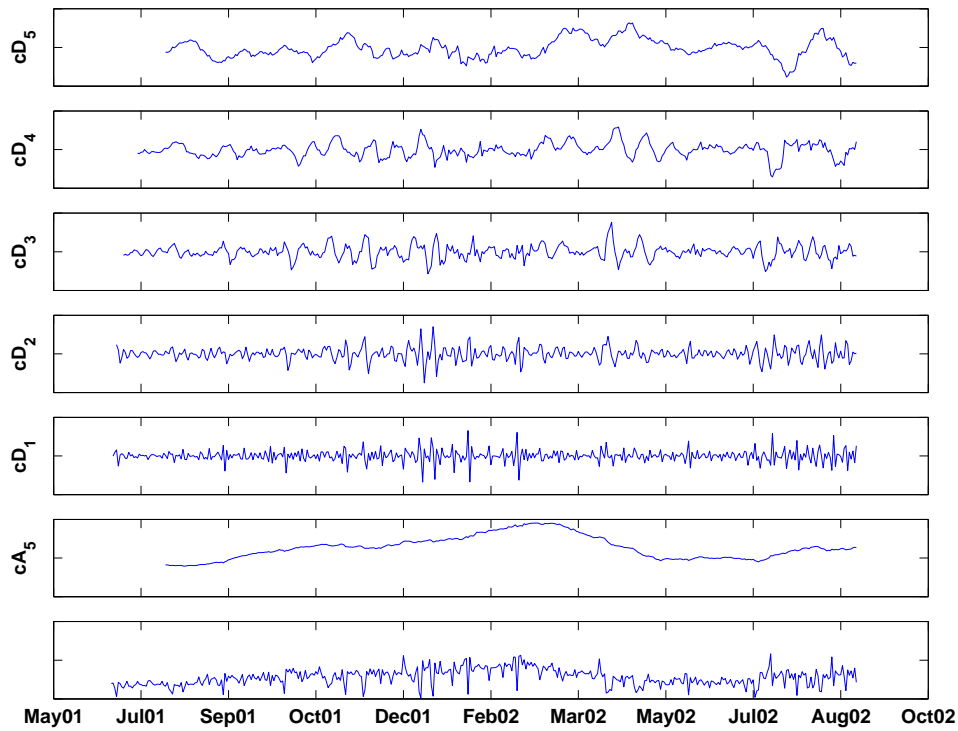


Figure 6: Decomposing the (deseasonalized) series of daily visits to military outpatient clinics into 5 levels using the “backward” Haar wavelet, without downsampling. The bottom panel displays the time series. Above it are the 5th approximation (cA<sub>5</sub>) and the the 5 levels of detail (cD<sub>5</sub>-cD<sub>1</sub>).

syndromic data we found that the detail coefficient series are well approximated by an autoregressive model of order 7 with zero-mean. A similar type of model was also been used by Goldenberg et al. (2002) to model series arising from a redundant wavelet transform of over-the-counter medication sales. Using an autoregressive model to forecast detail levels has also been suggested in other applications (e.g., Renaud et al., 2005; Xiao et al., 2005). The autoregressive model is used to forecast the coefficient (or its reconstruction) in the next point. In the context of monitoring, we need to specify how the AR parameters are estimated and how to derive forecast error measures. The following describes a roll-forward algorithm for monitoring the coefficients on a new day:

1. Using the phase I period, estimate the scale-specific AR(7) model coefficients  $(\theta_1, \dots, \theta_7)$  by using the first part of the phase I period, and the associated standard deviation of the forecast error  $(\sigma_{e_j})$  using the second part of the phase I period.
2. For the phase II data, forecast the next-day detail coefficient at level  $j$  ( $cD_j(t+1)$ ) using the estimated AR(7) model:

$$c\hat{D}_j(t+1) = \hat{\theta}_1 cD_j(t) + \hat{\theta}_2 cD_j(t-1) + \dots + \hat{\theta}_7 cD_j(t-6) \quad (2)$$

3. Using the estimated standard deviation of the forecast error at level  $j$  ( $\sigma_{D_j}$ ) create the control limits

$$c\hat{D}_j(t+1) \pm z_{1-\alpha/2} \cdot \hat{\sigma}_{e_j}. \quad (3)$$

4. Plot the next day coefficients on scale-level control charts with these control limits. If they exceeds the limits then an alarm is raised.

The approximation level captures the low-frequency information in the series, or the overall trend. This trend can be local, reflecting background noise and seasonality rather than an outbreak. It is therefore crucial to have a sense of what is a “no-outbreak” overall trend. When the outbreak of interest is associated with a bio-terrorist attack, we will most likely have enough “no-outbreak” data to learn about the underlying trend. However, to assess sampling variability of this trend, we need several no-outbreak seasons. For the case of “ordinary” disease outbreak detection, such as in our example, we have data from a single year, the trend appears to vary during the year, and this one year contains an

outbreak period. We therefore do not know whether to attribute the increasing trend to the outbreak or to seasonal increases in outpatient visits during those months. A solution would be to obtain more data for years with no such outbreaks (at least not during the same period), and to use those to establish a no-outbreak trend and its related sampling error. Of course, medical and epidemiological expertise should be sought in all cases in order to establish what is a reasonable level at different periods of the year.

Two methods for evaluating whether there is anomalous behavior using the scale-level forecasts are to add these forecasts to obtain a forecast for the actual new data point (as in Goldenberg et al., 2002). Alternatively, we can compare the actual vs. forecasted coefficient at each scale, and determine whether the actual coefficient is abnormal. In both cases we need estimates of the forecast error variability, which can be obtained from the forecast errors during phase I (the no-outbreak period). For the first approach of computing a forecast of the actual data, the standard deviation of forecast errors can be estimated from the forecasts generated during phase I and used to construct an upper control limit on the original time series. For the second approach, where scale-level forecasts are monitored directly, we estimate scale-specific standard deviations of the forecast errors from phase I (at scale  $j$  we estimate  $\hat{\sigma}_{e,j}$ ), and then construct scale-specific upper control limits on the coefficients at that scale. This is illustrated in the next section for our data.

In both cases, the algorithm is implemented in a roll-forward fashion: the scale-level models use the phase I data to forecast the next-day count or coefficients. This is then compared to the actual data or coefficients to determine whether an alarm should be raised. The new point is added to the phase I data, and the forecasting procedure is repeated with the additional data point.

Finally, an important factor for infiltrating new statistical methods into a non-statistician domain is software availability. Many software packages contain classic control charts, and they are also relatively easy to program. For this reason we make our wavelet-monitoring Matlab code freely available on the web. The code includes a program for performing retrospective DWT monitoring with Shewhart control limits, and a program for performing prospective SWT monitoring with Shewhart or AR-based control limits. In both cases it yields the control charts and lists the alarming dates. Outputs from this code

are shown in the next section.

### **3 Example: Monitoring Respiratory Complaints at Military Outpatient Clinics in Charleston, SC, between July 2001-August 2002**

We return with further detail to the syndromic data that we described in Section 1 and plotted in Figure 1. These are the daily number of respiratory-complaint visits to military outpatient clinics in Charleston, SC between July 2001 and August 2002. These are data that were used as part of the Bio-event Advanced Leading Indicator Recognition Technology (Bio-ALIRT) biosurveillance program by The Defense Advanced Research Projects Agency (DARPA), that took place between 2001-2004 (see Siegrist and Pavlin, 2004, for further details.)

The underlying assumption is that a disease outbreak will manifest itself earlier in this series than in diagnosis data, by showing an increase in daily visits. However, we do not know the pattern of increase: will it be a single-day spike, an exponential increase, or a step function? Furthermore, this time series exhibits a day-of-week effect, seasonality, dependence, and long-term dependence. Finally, a team of epidemiologists detected a gastrointestinal outbreak between Feb-19 and Mar-20 of 2002. Such an outbreak is associated with respiratory symptoms. However, as described in Section 1, the determination of this outbreak and its timing must be treated cautiously, as it relies on visual inspection of the data and the use of traditional monitoring tools.

We use this series to illustrate the performance of ordinary control charts (that are the standard tool in biosurveillance), and compare it to wavelet-based monitoring.

Common practice in biosurveillance is to use a standard control chart such as a CuSum or an EWMA chart to monitor the raw data (or its standardized version). For instance, Ivanov et al. (2003) use and EWMA to monitor daily hospital visits by children under the age of 5, who have respiratory and gastrointestinal complaints. Common false alarm rates that have been acceptable in biosurveillance are much higher than in engineering practices. The different algorithms in the Bio-ALIRT project

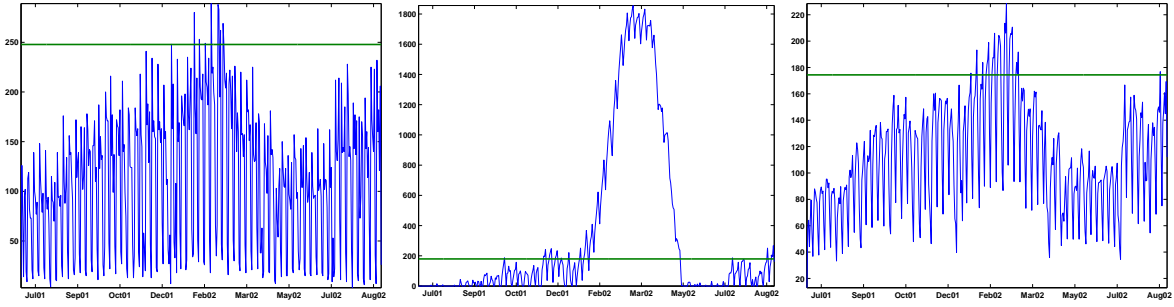


Figure 7: Shewhart x-chart (left), CuSum (center), and EWMA chart (right) applied to raw data.

report rates of a false alarm every 2-6 weeks. If for our 427-day data we aim for this rate, we expect approximately 10-30 days with false alarms. However, we hypothesize that the actual false alarm rate from control charts such as Shewhart, CuSum, and EWMA will be higher than the one we set it to because of the positive autocorrelation. We therefore set all alarm rates in the following to 1 in 100 days ( $\approx 4$  days in our time series).

Figure 7 illustrates the outcome of applying a one-sided Shewhart x-chart (with a 2.33-sigma upper limit), a one-sided CuSum (with  $k = 0.5, h = 2.84$ , using Sigmund's formula), and a one-sided EWMA chart (with  $\lambda = 0.4$  and a 2.33-sigma upper limit) directly to the raw data. The first 284 days (7/1/2001-12/31/2001) served as the phase I period and were used to compute the control limits. The Shewhart chart alarms on 9 days: 26-Dec-2001, [22,28]-Jan-2002, [4,11,19,20,22,25]-Feb-2002. All of these are weekdays, with the majority being Mondays. The CuSum alarms around 5 periods: a single alarm on Friday 12-Oct-2001, a month-long period with most days between 28-Nov to 28-Dec-2001, a 5-month period between 10-Jan-2002 and 10-May-2002, a single alarm on Friday 12-Jul-2002, and a week starting in 23-Aug-2002 (Friday to Friday). The EWMA alarms on every Friday between 11-Jan-2002 and 8-Mar-2002, with a few other weekdays in between. It alarms again on Friday, 23-Aug-2002. The common features to all charts are the lack of alarms on weekends, the alarms around the Jan-Feb. The last two charts also both alarm in late Aug-2002.

This brings up the important issue that syndromic data typically contain some periodicity. In our data there is an obvious 7-day periodicity. A close look shows that weekends (Sat-Sun) experience on average much less visits than weekdays, as expected. In addition, Mondays seem to incur the most visits. Applying a Shewhart, CuSum, or EWMA chart directly to periodic data, and especially using the

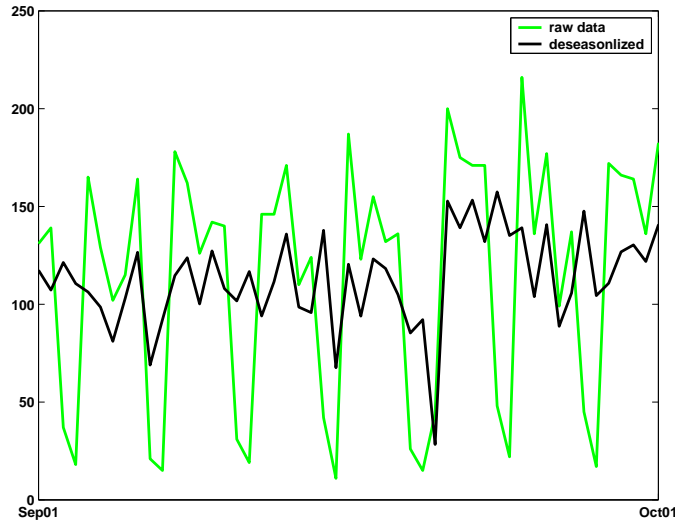


Figure 8: Raw counts (green) and deseasonalized (bold black) counts of daily visits during the month of September 2001.

periodic data to compute control limits, will obviously deteriorate their performance in two directions: the false alarm rate on days with high counts will increase, and the power of true detection on days with low counts will decrease. For our data we would therefore expect more false alarms on Mondays and low detection power on weekends. The control chart alarm days above indeed support this bias. Seasonality, therefore, should be accounted for before applying these control charts.

Forsberg et al. (2005)) apply a preliminary step of smoothing using a moving average with a window of 7 days to remove daily variation. However, this type of preliminary temporal averaging means that timeliness is sacrificed. Siegrist et al. (2005) describe a similar step and reported that “pre-filters using 2-7 day averages were also tried, and the detection delay defeated any gain from the data smoothing.” Other methods for accounting for day-of-week effects have been to model various days of the week as separate time series, and taking differences at a lag of 7.

A simple alternative that does not introduce temporal delays is to deseasonalize the series in a standard way. The deseasonalized series can then monitored using a standard control chart. Fig 8 shows the original and deseasonalized data for September 2001 (using multiplicative seasonal indexes based on moving averages).

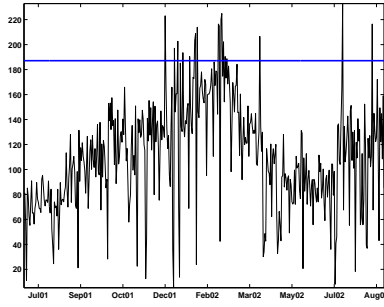


Figure 9: Shewhart chart with an upper 2.33-sigma limit applied to the deseasonalized data. The chart signals alarms on: 15,26,30-Dec-2001, 5,13,19-22-Jan-2002, and 16,17,19,20,22,24,26-Feb-2002, 6-Apr-2002, 13-Jul-2002, 17-Aug-2002.

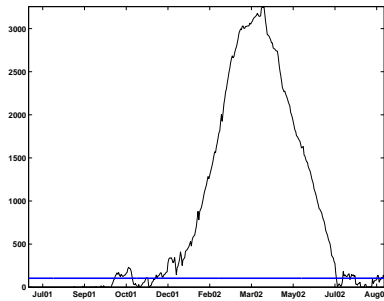


Figure 10: CuSum chart with  $k = 0.5$  and  $h = 2.84$  applied to the deseasonalized data. The chart signals alarms around 5 periods: 13-Oct-2001 to 2-Nov-2001; 19-20-Nov-2001, 30-Nov-2001 to 4-Jul-2002; 13-27-Jul-2002, and 23-31-Aug-2002.

### 3.1 Using standard control charts

Figures 9-11 display three standard (one-sided) control charts applied to the deseasonalized data: a 2.33-sigma Shewhart chart (9), a CuSum chart with  $k = 0.5$ ,  $h = 2.84$  (10) and an EWMA chart with  $\lambda = 0.4$  (11). As before, the first 284 days (7/1/2001-12/31/2001) were used to compute the control limits (phase I), and the alarm rate (under the *iid* assumption) was chosen for all cases to be 0.01.

Even with the reservation regarding the actual outbreaks in the data, it appears that the control charts are yielding too many false alarms relative to the 1% rate ( $\approx 4$  days) that they were set to. In all charts alarms are triggered around December-January, but the spread in dates is very large and not practical for deriving conclusions about outbreak dates. The CuSum and EWMA also alarm from

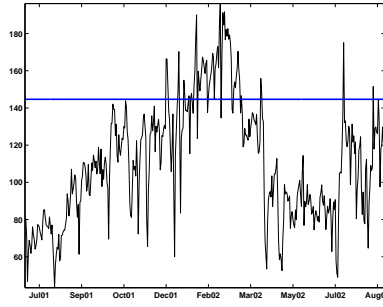


Figure 11: EWMA chart with  $\lambda = 0.4$  and an upper 2.33-sigma limit applied to the deseasonalized data. The chart signals alarms on: 15-17,29-30-Dec-2001, 5-6,11,13-14-Jan-2002; most days between 17-Jan-2002 and 14-Mar-2002; 6-7-Apr-2002, 13-Jul-2002, and 17-Aug-2002.

mid-July 2002 on. However, because of the reliance of these control charts on assumptions that are violated here, interpreting their signals is questionable. Finally, these charts are used in the same way for retrospective and prospective monitoring, because the daily statistic never relies on future values of the series.

## 3.2 Wavelet-based monitoring

We now use the DWT and SWT to decompose the deseasonalized data. We first decompose the first 284 days (7/1/2001-12/31/2001) in order to compute the control limits (phase I), as in the ordinary control chart case. In both cases we use the Haar wavelet and 5 levels of decomposition, because it is reasonable to compare a daily count with neighboring days up to a month (level 5 looks at a window of 32 days).

### 3.2.1 Retrospective monitoring

The left panel in Figure 12 shows the DWT of the (7-day deseasonalized) series, with a two-sided 2.58-sigma x-bar chart applied to each of the detail levels and a one-sided upper 2.33-sigma chart applied to the approximation level (because we are only interested in increases in the series). Different seasonal components can be seen at the different detail levels and possibly in the approximation. If at least one coefficient exceeds its thresholds on a certain day, it signals an alarm. Detail coefficients that are below

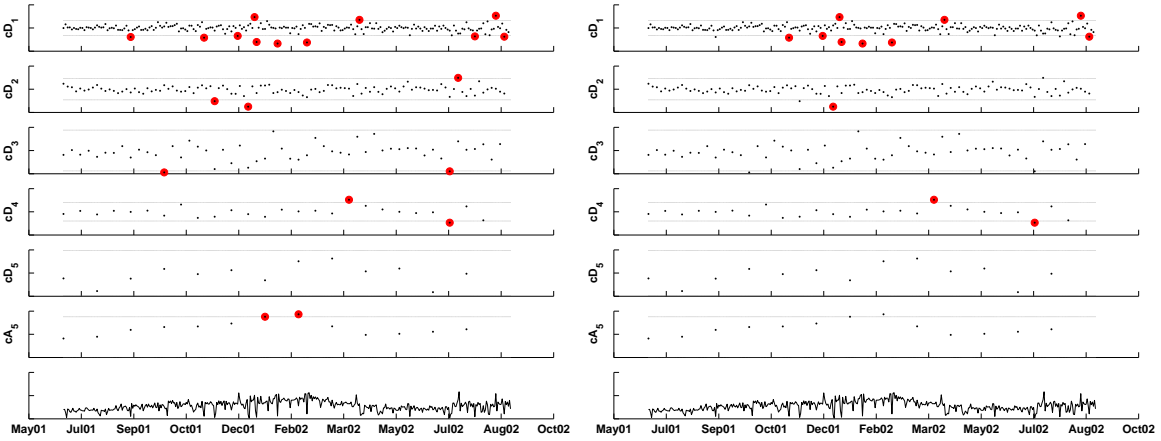


Figure 12: DWT of the time-series with 3-sigma  $\bar{x}$  charts applied at each scale. The left panel shows alarms based on coefficients that exceeded the thresholds. The right panel shows alarms after correcting for multiple testing.

the lower threshold indicate that the current period has lower counts than the next period, and the opposite holds for detail coefficients exceeding the upper threshold.

In this example almost all alarms are triggered by the finest detail coefficients indicating that the number of respiratory complaint visits on certain days (especially in Jan-Mar 2002) was high relative to their neighboring day (after accounting for the day-of-week effect). There is also alarming based on comparison with the counts two weeks previously. Notice that alarms are triggered also by the approximation level. However, it is unclear whether this indicates an abnormal increase in the series, or else it reflects “natural” seasonality in clinic visits during that period of the year. We discuss this ambiguity further in the next section. The right panel applies an FDR correction to the above results, to account for the multiple testing that occurs on many days. This results in reducing the alarm dates to 12 days: First on 12-Nov-2001, where the negative level-1 detail coefficient indicates that this Monday is much lower than the following Tuesday. The second period is 14,24,30-Dec-2001, 1-1-2002. Then, during the assumed 2002 outbreak period: 21-Jan, 18-Feb. And finally, a set of single-day alarms in 2002: 30-Mar, 9-Apr, 4-Jul, 17-Aug, and 25-Aug.

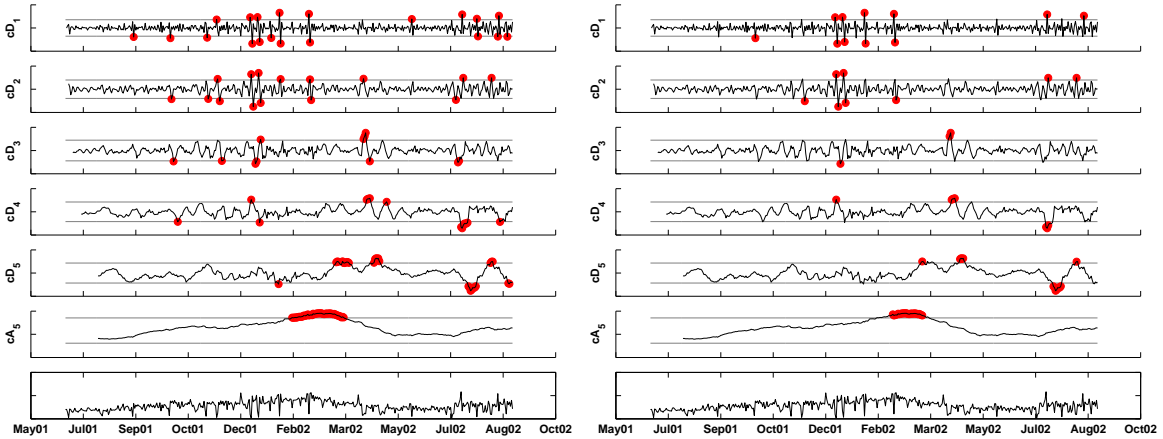


Figure 13: Backward-implemented SWT with 3-sigma  $\bar{x}$  charts applied at each scale. The left panel shows alarms based on coefficients that exceed the thresholds. The right panel shows alarms after correcting for multiple testing.

### 3.2.2 Prospective monitoring

As discussed earlier, DWT is better suited for retrospective surveillance than for prospective surveillance. Instead, we use SWT, which is an undecimated (non-downsampled) DWT, implemented in a “backward” fashion, so that coefficients at time  $t$  are computed only using data before time  $t$ . This is mathematically equivalent to a role-forward algorithm that starts after Phase I ends, and performs this SWT version every day until Aug-31-2002.

Figure 13 shows the result of decomposing the data by SWT and applying a 3-sigma  $\bar{x}$  chart to each scale. It can be seen that, as expected, the coefficients within each scale are no longer approximately decorrelated, as are coefficients across scales. The FDR correction is used to correct for the correlation across scales, and the right panel shows the remaining alarms after applying the FDR correction. Even after the FDR correction we get alarms for long periods of time: the first is between 17-Jan-2002 and 25-Feb-2002, and then we get three single day alarms on 22-Jun-2002, 30-Jun-2002, and 13-Aug-2002. These results are not trustworthy because of the heavy correlation within each scale. We therefore used AR(7) models to forecast the next-day detail coefficients. Using the phase I period, we estimated the scale-specific model coefficients using the first 92 days and the associated standard deviation of the forecast errors from the next 92 days. The estimates are given in Table 1. Residual plots indicate

|                      | $cD_1$  | $cD_2$  | $cD_3$  | $cD_4$  | $cD_5$  |
|----------------------|---------|---------|---------|---------|---------|
| $\hat{\theta}_1$     | -0.8746 | 0.7367  | 0.8580  | 1.0040  | 0.9760  |
| $\hat{\theta}_2$     | -0.6520 | -1.1321 | 0.0177  | -0.0990 | 0.2617  |
| $\hat{\theta}_3$     | -0.4758 | 0.8163  | -0.1059 | 0.0637  | -0.0174 |
| $\hat{\theta}_4$     | -0.3598 | -0.8944 | -0.6849 | -0.1315 | -0.1596 |
| $\hat{\theta}_5$     | -0.3205 | 0.4981  | 0.6003  | 0.0631  | -0.1195 |
| $\hat{\theta}_6$     | -0.2248 | -0.3905 | -0.0705 | 0.0285  | -0.1017 |
| $\hat{\theta}_7$     | -0.1610 | 0.1352  | -0.0986 | -0.1413 | 0.1209  |
| $\hat{\sigma}_{e_j}$ | 23.1    | 15.2    | 10.6    | 7.4     | 4.0     |

Table 1: Estimated coefficients and standard deviation of forecast errors for scale-level AR(7) models, based on the first 184 days.

that the model residuals are approximately normally distributed, and thus we can use equation (3) to construct control limits for the coefficient series. This is shown in the top 5 panels of Figure 14.

Monitoring the approximation level is challenging in our example because we have data only on a single year, and it supposedly contains an outbreak of the type we are interested in detecting. It is therefore unknown whether the increasing trend during the Jan-Mar-2002 period is due to the outbreak, or to other unrelated reasons. In this case it is necessary to have more data on the Jan-Mar season where it can be asserted that no outbreak of this type has occurred. For sake of illustration, let us assume that past data indicate that there is no increasing trend during Jan-Mar in such syndromic data. In that case we can use a chart such as an EWMA to monitor the approximation coefficient series (because of the autocorrelation). The right panel of Figure 14 shows the result of applying an FDR correction for the multiple testing. So here we are accounting for both correlation within scales and across scales. The remaining alarms occur twice in the finest detail coefficients (21-Jan-2002, 18-Feb-2002), and the remaining are all in the approximation (30-12-2001, and 9-Jan-2002 to 9-Apr-2002.) Recall that this is based on an assumption that the trend increase during these months does not reflect normal behavior of this series. In the presence of data on more years we would have been able to assess this better and fit a model to the approximation series based on the trend information.

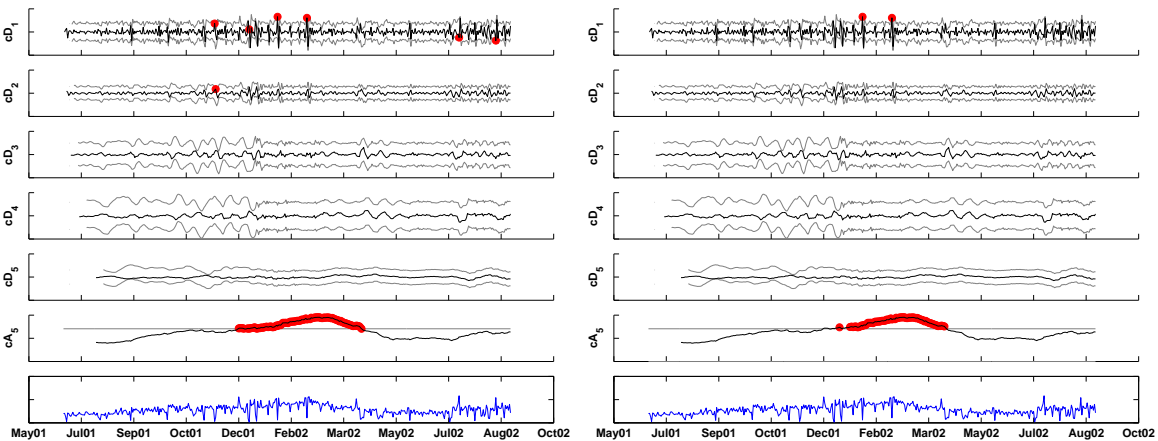


Figure 14: Monitoring each scale separately to account for autocorrelation: the detail coefficients (top 5 panels) are compared to control limits based on an AR(7) model. The approximation coefficients (6th panel) are monitored by an EWMA chart. The bottom panel is the deseasonalized series. Left panel displays alarms based on exceeding thresholds. Right panel displays alarms after FDR correction.

Finally, we can see that the prospective monitoring gives different (and in this case, less) alarms compared to the retrospective monitoring. This is because the prospective algorithm assess coefficients by comparing them only to their past, whereas the retrospective algorithm compares each point to its past and future.

## 4 Conclusions and Future Directions

The goal of this paper is to introduce the important area of modern biosurveillance and the challenges that it poses to traditional statistical monitoring. There are currently not many statisticians involved, and there is a pressing need to develop improved biosurveillance systems. From a research point of view, there are opportunities for developing statistical methodology for improving the development and evaluation of biosurveillance systems. Three components make biosurveillance challenging. First, Like many other fields, the advancement of technology has lead to data that are more complex than those collected a century ago. More frequent and diverse data mean that the classical assumptions of sample-to-sample independence and stationarity tend not to be met. This challenge is not specific

to bio-surveillance and is apparent in chemical processes, geo-physical data, etc. Second, modern biosurveillance systems monitor syndromic and non-traditional data that are assumed to contain an earlier signature of an outbreak than actual diagnosis data. However, this earliness comes at the cost of a weaker signal of the outbreak compared to actual diagnosis data. But they . This means that detecting the signal commands sensitive control methods that are timely. Third, since there is nearly no syndromic data that contain bio-terrorist outbreaks, it is unknown how such an outbreak would manifest itself in the data. This means that classic control charts that are specialized in detecting a particular type of pattern (e.g., a single spike, an exponential increase, or a linear trend), are risky. Methods that are “general detectors” reduce the risk by “diversifying” the detection to a wider number of patterns. This situation also occurs in other fields, where the outbreak nature might not be known apriori (e.g. in forecasting storms). Finally, the issue of lack of “golden standards” is a major challenge. The problem is that determining whether an outbreak of interest is contained in the data is not straightforward. The current practice is for a team of medical and epidemiologists to eyeball a few series of syndromic data in order to determine whether and when an outbreak started. This clearly results in diverting the tuning and development of monitoring tools to detecting what the team sees, rather than actual outbreaks in the data. Furthermore, it greatly complicates the evaluation of algorithms and their performance. The implications of not having golden standards are: 1. When the goal is to detect natural outbreaks and we do not know whether and when exactly in the data there are such outbreaks, it is hard to assess what is a phase I in order to estimate process in-control parameters and to establish control limits. 2. When the goal is to detect outbreaks associated with bio-terrorist attacks we can (luckily) assume that the data are clean of such attacks. However, the presence of natural outbreaks in the data create more background noise that is hard to model if it is not specified as an outbreak. One approach has been to try and seed the data with outbreaks (e.g., Goldenberg et al. (2002); Stoto et al. (2005)). This avoids the lack of data with a certain outbreak signature in it, but we still have the problem of determining whether other outbreaks occur in the data. Furthermore, outbreak simulation is challenging because we do not know what the pattern will look like. If we knew, we would design a good monitoring tool to detect that pattern. When simulating a certain type of outbreak, we automatically give an advantage to

some methods which can be specified apriori (e.g. a CuSum for detecting a small step function change).

Current biosurveillance relies on classical control charts such as the CuSum and EWMA. For the reasons mentioned above we believe that these tools are not always adequate for the purpose and requirements of biosurveillance. Shmueli and Fienberg (2005) survey advanced monitoring methods in different fields and assess their potential for biosurveillance. One of these is based on wavelets. In this work we introduce a general approach based on wavelets. The main idea is to decompose a series into a time-frequency domain and then monitor the different scale for abnormalities. Wavelet-based methods have the advantage of making less assumptions about the data (regarding independence and stationarity), they are “general detectors” in the sense of not being specified to a particular abnormality pattern, and they are computationally efficient. There are different ways to use wavelet decompositions for monitoring. We discussed different options, their advantages and limitations, and what they would be suitable for. There are most likely other ways to integrate wavelet decompositions with monitoring, which could lead to improved methods. Finally, a generalization to multivariate monitoring would be a large step in biosurveillance monitoring. Ordinary multivariate monitoring tools such as the Hotelling- $T^2$  and multivariate-CuSum charts are limited by the same assumptions as their univariate counterparts. Wavelet-based generalizations would therefore be a potential powerful tool for biosurveillance.

## Acknowledgements

The author thanks Dr. Howard Burkom from the Johns Hopkins Applied Physics Lab, Prof. Paul Kvam from the School of Industrial and Systems Engineering at Georgia Tech, and Henry Rolka from the Centers for Disease Control for helpful discussions.

## References

- Abramovich, F., Bailey, T., and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician - Journal of the Royal Statistical Society, Ser. D*.
- Aradhye, H. B., Bakshi, B. R., Strauss, R. A., and F, D. J. (2003). Multiscale statistical process control using wavelets - theoretical analysis and properties. *AIChE Journal*, 49(4):939–958.
- Aussem, A. and Murtagh, F. (1997). Combining neural network forecasts on wavelet-transformed time series. *Connection Science*, 9:113–121.
- Bakshi, B. R. (1998). Multiscale pca with application to multivariate statistical process monitoring. *AIChE Journal*, 44:1596–1610.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57:289–300.
- Box, G. and Luceo, A. (1997). *Statistical Control: By Monitoring and Feedback Adjustment*. Wiley-Interscience, 1st edition.
- Burkom, H. S., Elbert, Y., Feldman, A., and Lin, J. (2004). Role of data aggregation in biosurveillance detection strategies with applications from essence. *Morbidity and Mortality Weekly Report (MMWR)*, 53 (suppl):67–73.
- Fienberg, S. E. and Shmueli, G. (2005). Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine (to appear)*.
- Forsberg, L., Jeffery, C., Ozonoff, A., and Pagano, M. (2005). *Statistical Methods in Counter Terrorism*, chapter A Spatio-temporal Analysis of Syndromic Data for Biosurveillance. ASA-SIAM.
- Gencay, R., Selcuk, F., and Whitcher, B. (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. Academic Press.
- Goldenberg, A., Shmueli, G., Caruana, R. A., and Fienberg, S. E. (2002). Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. *Proceeding of the National Academy of Sciences*, 99:5237–5240.

- Heffernan, R., Mostashari, F., Das, D., Besculides, M., Rodriguez, C., Greenko, J., Steiner-Sichel, L., Balter, S., Karpati, A., Thomas, P., Phillips, M., Ackelsberg, J., Lee, E., Leng, J., Hartman, J., Metzger, K., Rosselli, R., and Weiss, D. (2004). System descriptions new york city syndromic surveillance systems. *Morbidity and Mortality Report Weekly (MMWR)*, 53 (suppl):23–27.
- Ivanov, O., Gesteland, P. H., Hogan, W., Mundorff, M. B., and Wagner, M. M. (2003). Detection of pediatric respiratory and gastrointestinal outbreaks from free-text chief complaints. In *AMIA Annu Symp*, number 318-322.
- Ku, W., Storer, R. H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30:179–196.
- Mastrangelo, C. M. and Montgomery, D. C. (1995). Spc with correlated observations for the chemical and process industries. *Quality and Reliability Engineering International*, 11(2):78–89.
- Percival, D. and Walden, A. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, U.K.
- Reis, B. and Mandl, K. (2003). Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making*, 3(2).
- Renaud, O., Starck, J.-L., and Murtagh, F. (2005). Wavelet-based combined signal filtering and prediction. *IEEE Transactions SMC, Part B*, (in press).
- Runger, G. C. and Willemain, T. R. (1995). Model-based and model-free control of autocorrelated processes. *Journal of Quality Technology*, 27:283292.
- Serfling, R. E. (1963). Methods for current statistical analysis fo excess pneumonia-influenza deaths. *Public Health Reports*, 78:494–506.
- Shmueli, G. and Fienberg, S. E. (2005). *Statistical Methods in Counter Terrorism*, chapter Current and potential statistical methods for monitoring multiple data streams for bio-surveillance. ASA-SIAM.
- Siegrist, D., McClellan, G., Campbell, M., Foster, V., Burkom, H., Hogan, W., Cheng, K., Buckeridge,

- D., Pavlin, J., and Kress, A. (2005). Evaluation of algorithms for outbreak detection using clinical data from five u.s. cities. Technical report, DARPA Bio-ALIRT Program.
- Siegrist, D. and Pavlin, J. (2004). Bio-alert biosurveillance detection algorithm evaluation. *Morbidity and Mortality Weekly Reports (MMWR)*, 53 (suppl):152–158.
- Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society B*.
- Stoto, M., Fricker, R. D., Jain, A., Davies-Cole, J. O., Glymph, C., Kidane, G., Lum, G., Jones, L., Dehan, K., and Yuan, C. (2005). *Statistical Methods in Counter Terrorism*, chapter Evaluating Statistical Methods for Syndromic Surveillance. ASA-SIAM.
- Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika*, 82:385–397.
- Xiao, H., Sun, H., Ran, B., and Oh, Y. (2005). The fuzzy-neural network traffic prediction framework with wavelet decomposition. *Transportation Research Record, Journal of the Transportation Research Board*, page upcoming.
- Zhang, J., Tsui, F., Wagner, M., and Hogan, W. (2003). Detection of outbreaks from time series data using wavelet transform. In *AMIA Annual Symposium Proceedings*, pages 748–752.