

## Chapter 8

# ALGORITHM COMBINATION FOR IMPROVED PERFORMANCE IN BIOSURVEILLANCE

## *Univariate Monitoring*

INBAL YAHAV<sup>1,\*</sup>, THOMAS LOTZE<sup>2</sup>, and GALIT SHMUELI<sup>1</sup>

### CHAPTER OVERVIEW

This chapter proposes an enhancement to currently used algorithms for monitoring daily counts of pre-diagnostic data. Rather than use a single algorithm or apply multiple algorithms simultaneously, our approach is based on ensembles of algorithms. The ensembles lead to better performance in terms of higher true alert rates for a given false alert rate. Combinations can be employed at the data preprocessing step and/or at the monitoring step. We discuss the advantages of such an approach and illustrate its usefulness using authentic modern biosurveillance data.

**Keywords:** Control charts; Monitoring; Ensemble methods; Optimization; Pre-diagnostic data

### 1. INTRODUCTION

In 1918 one of the deadliest influenza pandemics in history erupted, called the Spanish Flu. Approximately 20–40% of the worldwide population fell ill and over 50 million people died. Outbreaks followed shipping routes

---

<sup>1\*</sup> *Department of Decision, Operations & Information Technologies, Robert H Smith School of Business, University of Maryland, College Park, MD 20742, USA*

<sup>2</sup> *Applied Mathematics & Scientific Computation Program, University of Maryland, College Park, MD 20742, USA*

from North America through Europe, Asia, Africa, Brazil and the South Pacific. The pandemic reached its peak after 5–6 months. Nearly 40 years later, in February 1957, the Asian influenza pandemic erupted in the Far East. Unlike the Spanish Flu, the Asian influenza pandemic virus was quickly identified and vaccines were available 6 months later. Approximately two million people died in this outbreak (compared to the 50 million in the Spanish Flu). Other known outbreaks in history, such as the Hong Kong Flu (1968–1969), the Avian Flu (1997) and SARS (2003) also resulted in high death tolls over the years. Unfortunately the threat of new pandemic outbreaks is still looming.

A major goal of public health is to figure out whether and how transmission of diseases can be diminished. Researchers at the Center for Humanitarian Logistics at Georgia Tech have shown that pandemic outbreak effects can be greatly reduced if quarantine is imposed at the early stages of the disease.<sup>1</sup> The U.S. Centers for Disease Control & Prevention (CDC) lay out guidelines and strategies for reducing disease transmission, including use of personal protective equipment (e.g., masks and gloves), hand hygiene, and safe work practices. The CDC also recommends actions to be taken during the earliest stage of a pandemic, when the first potential cases or disease clusters are detected. These include individual-level containment measures such as patient isolation and identification, monitoring, and quarantine of contacts.<sup>2</sup>

The early detection of disease outbreaks therefore plays a major role in preventing disease transmission and reducing the size of the affected population. In modern biosurveillance a wide range of pre-diagnostic and diagnostic daily counts are monitored for the purpose of alerting public health officials when there is early evidence of a disease outbreak. This is in contrast to traditional biosurveillance, where only diagnostic measures (such as mortality and lab reports) are examined, usually locally, and at aggregation levels such as weekly, monthly, or annually. Moreover, in modern biosurveillance the goal is prospective while traditional biosurveillance is more retrospective in nature. Although the tasks and data types and structures differ widely between traditional and modern biosurveillance, most monitoring algorithms have been migrated from traditional to modern systems. The result is that current modern biosurveillance detection methods suffer from multiple statistical and practical limitations that greatly deteriorate their ability to achieve their intended purpose. For a general overview of the statistical challenges that arise in biosurveillance see Shmueli and Burkom (2008). In particular, there is often a mismatch between the types of

---

<sup>1</sup> <http://www.tli.gatech.edu/research/humanitarian/projects.php>

<sup>2</sup> <http://www.hhs.gov/pandemicflu/plan/appendixf.html>.

algorithms being used and the data structure of modern biosurveillance data. This means that the assumptions behind those algorithms with regards to baseline behavior and outbreak nature are often violated in the modern biosurveillance context. Another important problem is that of multiplicity: when monitoring multiple data sources with multiple streams, using multiple monitoring algorithms, the unavoidable result is a large inflation in false alarm rate. In the case of biosurveillance the implication of excess false alerts outweighs the benefit of an early detection, as the practical result is that public health users ignore alerts altogether. An excess of false alerts usually leads to the ignoring of true alerts, and potentially to deletion of the alerting system altogether. An example is the rocket-alert system in the city of Ashkelon, Israel that was recently disconnected because of five false alarms in April 2008 which led to panic. Thus, when a rocket fell on a shopping mall in Ashkelon the following month, there was no early warning.<sup>3</sup>

In this chapter we focus on a solution to two important problems: that of multiplicity in monitoring algorithms and the unknown nature of the outbreak signature. We show that by combining results from multiple algorithms in a way that controls the overall false alert rate, we can actually improve overall performance. The remainder of the chapter is organized as follows. Section 2 describes control charts in general and the limitations of applying them directly to raw modern biosurveillance data. It then describes a pre-processing step that is needed before applying control charts. In Sect. 3 we describe an authentic modern biosurveillance dataset and the simulation of outbreak signatures. Section 4 introduces the notion of model combinations in terms of combining residuals and combining control chart output. Section 5 applies the different combination methods to our data, and we display results showing the improvement in detection performance due to method combination. Section 6 summarizes the main points and results and describes potential enhancements.

## 2. CONTROL CHARTS AND BIOSURVEILLANCE

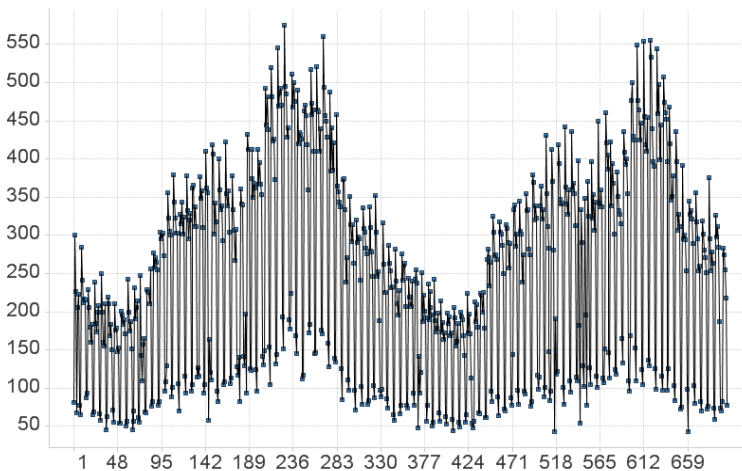
Control charts (also referred to as monitoring charts) are used to monitor a process for some quality parameter in order to detect anomalies from desired behavior. In the context of modern biosurveillance, control charts are used to monitor aggregated daily counts of individual healthcare seeking behavior (such as daily arrivals to emergency departments or medication sales), for the purpose of early detection of shifts from expected baseline behavior. Three control charts are commonly used to monitor such pre-

---

<sup>3</sup> <http://www.haaretz.com/hasen/spages/983479.html>, accessed May23, 2008.

diagnostic daily data, and are implemented (with some variations) in the three main national biosurveillance systems in the U.S.: BioSense (by CDC), ESSENCE (by DoD), and RODS. The three charts are Shewhart, Cumulative Sum (CuSum) and Exponential Weighted Moving Average (EWMA) charts. These control charts are described in detail in Sect. 2.1.

Using control charts to monitor biosurveillance data has two major drawbacks. First, control charts assume that the monitored statistics follow an independent and identically-distributed (iid) normal distribution with constant mean and variance. Daily pre-diagnostic counts usually fail to meet this assumption. In reality time series of such daily counts often contain seasonal patterns, day-of-week effects, and holiday effects (see [Figure 8-1](#) for illustration). Monitoring such data therefore requires an initial processing step where such explainable patterns are removed. Such methods are described in Sect. 2.2. For illustration, compare [Figures 8-1](#) and [8-2](#) that show a series of daily military clinic visits before and after preprocessing. One explainable pattern that is removed is the day-of-week effect, which is clearly visible in [Figure 8-1](#), but absent from [Figure 8-2](#).



*Figure 8-1.* Raw series of number of daily military clinic visits with respiratory complaints.

The second challenge of applying control charts in the modern biosurveillance context is that each type of chart is most efficient at capturing a specific outbreak signature (Box and Luceno, 1997). Yet, in the context of biosurveillance the outbreak signature is unknown, and in fact the goal is to detect a wide range of signatures for a variety of disease outbreaks, contagious and non-contagious, both natural and bioterror-related. It is therefore unclear which method should be used to detect such a wide range of unspecified anomalies.

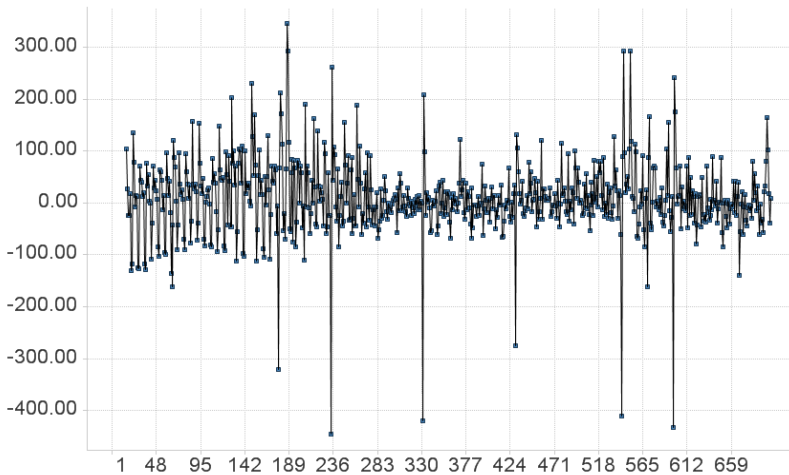


Figure 8-2. Daily military clinic visits series after removing explainable patterns.

## 2.1 Control Chart Overview

We briefly describe the most popular control charts in statistical quality control, which are widely used in modern biosurveillance systems:

**Shewhart.** The Shewhart chart is the most basic control chart. A daily sample statistic (such as a mean, proportion, or count) is compared against upper and/or lower control limits (UCL and LCL), and if the limit(s) are exceeded, an alarm is raised. The control limits are typically set as a multiple of standard deviations of the statistic from the target value (Montgomery, 1997). It is most efficient at detecting medium to large spike-type outbreaks.

**CuSum.** Cumulative Sum (CuSum) control charts monitor cumulative sums of the deviations of the sample statistic from the target value. CuSum is known to be efficient in detecting small step-function type changes in the target value (Box and Luceno, 1997).

**EWMA.** The Exponentially Weighted Moving Average (EWMA) chart monitors a weighted average of the sample statistics with exponentially decaying weights (NIST/SEMATECH Handbook). It is most efficient at detecting exponential changes in the target value and is widely used for detecting small sustainable changes in the target value.

Table 8-1 summarizes for each of the three charts its monitoring statistic (denoted  $Shewhart_t$ ,  $EWMA_t$  and  $CuSum_t$ ), the upper control limit (UCL) for alerting, the parameter value that yields a theoretical 5% false alert rate, and a binary output indicator that indicates whether an alert was triggered on day  $t$  (1) or not (0).  $Y_t$  denotes the raw daily count on day  $t$ . We consider one-sided control charts where an alert is triggered only when there is indication

of an increase in mean (i.e., when the monitoring statistic exceeds the UCL). This is because only increases are meaningful in the context of healthcare seeking counts.

Table 8-1. Features of three main control charts.

	Shewhart	EWMA	CuSum
Monitored	Shewhart <sub>t</sub> = $Y_t$	EWMA <sub>t</sub> =	CuSum <sub>t</sub> =
Statistic		$\lambda Y_t + (1 - \lambda)EWMA_{t-1}$	Max (0, CuSum <sub>t-1</sub> + $Y_t - \sigma/2$ )
UCL	UCL = $\mu + k \times \sigma$	UCL = EWMA <sub>0</sub> + $k \times \sigma$ $s^2 = \lambda/(2 - \lambda) \times \sigma^2$	UCL = $\mu + h \times \sigma$
Theoretical 5% Threshold	k = 1.5	k = 1.5	h = 2.5
Output	$S_t = \text{if}$ [Shewhart <sub>t</sub> >UCL]	$E_t = \text{if}$ [EWMA <sub>t</sub> >UCL]	$C_t = \text{if}$ [CuSum <sub>t</sub> >UCL]

## 2.2 Preprocessing Methods

There are a variety of methods for removing explainable patterns from time series. Methods generally are either model-based or data-driven. Model-based methods remove a pattern by directly modeling the pattern via some specification. An example is a linear regression model with day-of-week indicators. Data-driven methods either suppress certain patterns (e.g., differencing at a certain lag) or “learn” patterns from the data (e.g., exponential smoothing). In the following we describe three methods that have been shown to be effective in removing the types of explainable effects that are often exhibited in pre-diagnostic daily count series (day-of-week, holiday, seasonal, and autocorrelation). For a more detailed discussion of preprocessing methods see Lotze et al. (2008) and Lotze and Shmueli (2008). In the following we describe three methods that produce next-day forecasts. The forecasts are then subtracted from the actual counts to produce residuals.

**Method 1:** Holt-Winters exponential smoothing, using smoothing parameter values  $\alpha = 0.4$ ,  $\beta = 0$ , and  $\gamma = 0.15$  as suggested in Burkom et al. (2007). In addition, we do not update the forecasting equation if the percentage difference between the actual and fitted values is greater than 0.5.

**Method 2:** 7-day differencing (residuals are equal to the difference between the values of the current day and the same day 1 week previous). Equivalently, forecasts are obtained by using the values from 1 week previous.

**Method 3:** Linear regression of daily counts, using as covariates sine and cosine yearly seasonal terms, six day-of-week dummy variables, and a linear trend term. Only data in the first year are used for parameter estimation (training data).

### 3. DATA AND OUTBREAKS

#### 3.1 Data Description

Our data is a subset of the dataset used in the BioALIRT program conducted by the U.S. Defense Advanced Research Projects Agency (DARPA) (Siegrist and Pavlin, 2004). The data include six series from a single city, where three of the series are indicators of respiratory symptoms and the other three are indicators of gastrointestinal symptoms. The series come from three different data sources: military clinic visits, filled military prescriptions, and civilian physician office visits. Figures 8-3 and 8-4 display the six series of daily counts over a period of nearly 2 years. We illustrate the methods throughout this chapter using the series of respiratory symptoms collected from military clinic visits (top panel).

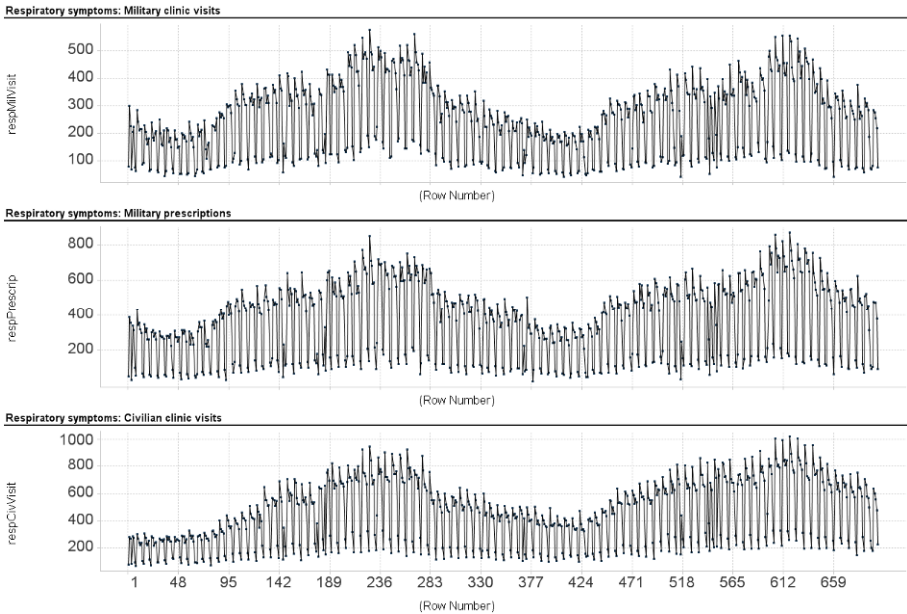


Figure 8-3. Daily counts of military clinic visits (*top*), military filled prescriptions (*middle*) and civilian clinic visits (*bottom*), all respiratory-related.

#### 3.2 Outbreak Signatures

Before preprocessing the raw data, we inject into the raw data outbreak signatures. The insertion into the raw data means that we assume that effects such as day-of-week and holidays will also impact the additional counts due to an outbreak. We simulate two different outbreak signature shapes: a

single-day spike and a multiple-day lognormal progression. We set the size of the affected population to be proportional to the variance of the data series (Lotze et al., 2007).

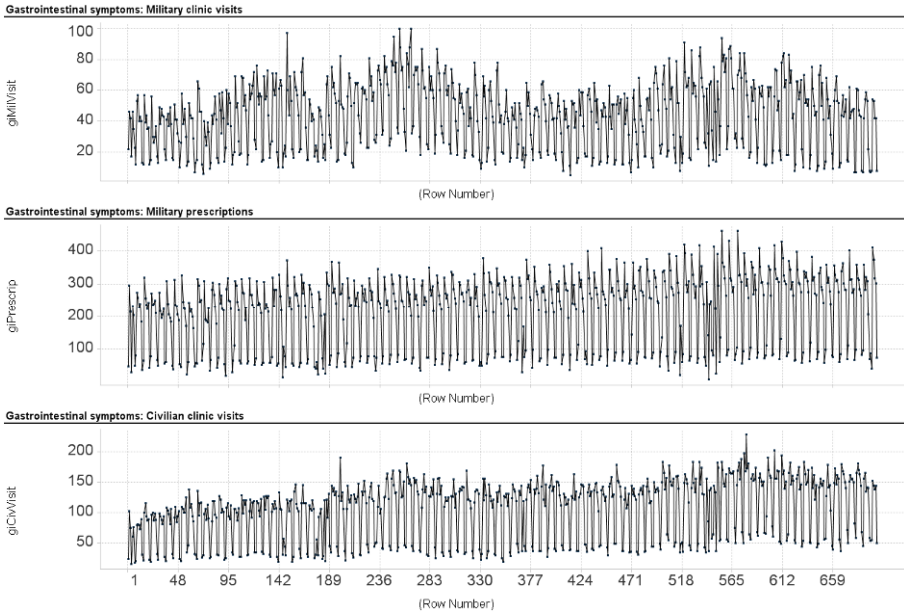


Figure 8-4. Daily counts of military clinic visits (*top*), military filled prescriptions (*middle*) and civilian clinic visits (*bottom*), all gastrointestinal-related.

For the single-day spike, we consider small to medium spike sizes, because biosurveillance systems are designed to detect early, more subtle indications of a disease outbreak. We also consider a lognormal progression signature, because incubation periods have been shown to follow a lognormal distribution with parameters dependent on the disease agent and route of infection. In order to generate a trimmed lognormal signature (Burkom, 2003), we set the mean of the lognormal distribution to 2.5 and the standard deviation to 1. We trim 30% of the long tail, limiting the outbreak horizon to approximately 20 days. This choice of parameters results in a gradually increasing outbreak with a slow fading rate (long tail). Figure 8-5 illustrates the process of injecting a lognormal outbreak into the raw data.

## 4. COMBINATION MODELS

We consider the problem of linearly combining residuals and/or control chart output vectors for improving the performance of automated bio-surveillance algorithms. In order to better evaluate the contribution of each



of the two levels of combination, we first examine residual combinations and control chart combinations separately: when combining residuals from different preprocessing techniques, we use a single control chart (see Figure 8-6); when combining control chart outputs we use a single preprocessing technique (see Figure 8-7). We then examine the additional improvement in performance from optimizing the complete process (combining both residuals and control charts).

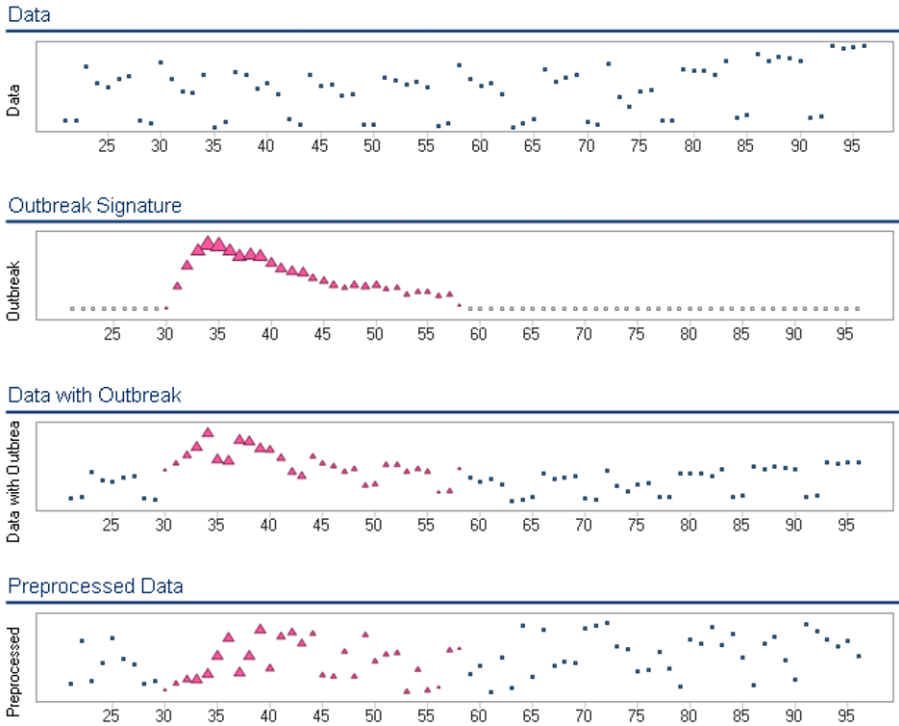


Figure 8-5. Injecting a lognormal outbreak signature into raw data, and preprocessing the resulting series.

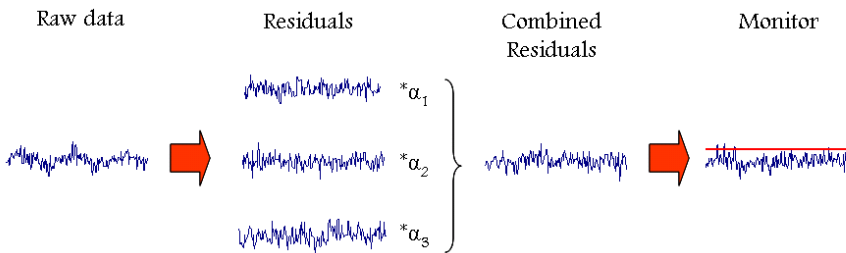


Figure 8-6. Combining residuals.

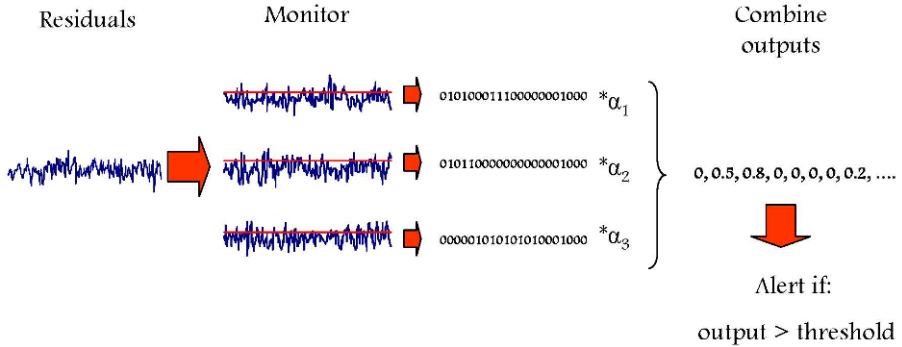


Figure 8-7. Combining control chart outputs.

We assume that the data are associated with a label vector  $O_t$ , which denotes whether there is an actual outbreak at day  $t$ . We further assume a sufficient amount of training data. The labeled vector and sufficient training data are essential when seeking an optimal combination that increases the true alert rate while maintaining a manageable false alert rate.

## 4.1 Residual Combination

The idea of using an ensemble is inspired by machine learning classifier techniques, which have produced improved classification by combining multiple classifiers. We used a simple method as our main combination method: a linear combination of next-day forecasts that minimizes the mean squared errors of past data. The coefficients for this linear combination can be determined using linear regression, with the forecasters as predictors and the actual value as the dependent variable. We also compared combinations using a day-of-week modification: residual combinations are optimized separately using only past data from the same day of the week.

## 4.2 Control Chart Combination

In this section, we assume that the raw data have undergone a pre-processing step for removing explainable patterns. Thus, the input into the control charts is a series of residuals. We consider the three monitoring charts described in Sect. 2: Shewhart, EWMA and CuSum. We construct a linear combination of the monitoring binary output for the purpose of maximizing the true alert rate, while constraining the false alert rate to be below a specific threshold. This formulation yields the following mixed integer programming optimization problem:

$$\max \sum_{t=1}^n TA_t$$

s.t.

$$(Bin:) \quad FA_t, MA_t \in \{0,1\}$$

$$(FA:) \quad (w_S \times S_t + w_E \times E_t + w_C \times C_t) \times (1 - O_t) - T < FA_t$$

$$(TA1) \quad [(w_S \times S_t + w_E \times E_t + w_C \times C_t) - T] \times O_t \leq TA_t \times O_t$$

$$(TA2) \quad TA_t \times O_t \leq (w_S \times S_t + w_E \times E_t + w_C \times C_t) \times O_t$$

$$(FA\_sum:) \quad \sum_{t=1}^n FA_t < \alpha \times n$$

where  $w_i$  is the weight of control chart  $i$ , and  $FA_t$  ( $TA_t$ ) is an indicator for a false (true) alert on day  $t$ . The constraints can be interpreted as follows:

*(Bin)* restricts the false alert ( $FA$ ) and true alert ( $TA$ ) indicators on day  $t$  to be binary.

*(FA)* is a set of  $n$  (training horizon) constraints that determine whether the combined output  $w_S \times S_t + w_E \times E_t + w_C \times C_t$  yields a false alert on day  $t$ :

If there is an outbreak on day  $t$ , then  $1 - O_t = 0$  and the constraint is satisfied.

Otherwise ( $1 - O_t = 1$ ) we compare the combined output with the threshold  $T = 1$ . If the combined output is greater than the threshold, we set  $FA_t$  to 1.

Similarly *(TA1 and TA2)* is a set of  $2n$  constraints that determine whether the combined output  $w_S \times S_t + w_E \times E_t + w_C \times C_t$  yields a true alert on day  $t$ .

Finally, we set the false alert rate to be less than  $\alpha$  (*FA\_sum*).

## 5. EMPIRICAL STUDY AND RESULTS

In this section we describe the results obtained from applying the combination methods to authentic pre-diagnostic data with simulated outbreaks. We start by describing the experimental design and then evaluate the methods' performance.

### 5.1 Experiment Design

We inject into the raw series 100 outbreak signatures, in random locations (every 10 weeks on average). Each outbreak signature is a spike of size  $0.5 \times \sigma$  ( $\sim 60$  cases), with probability 0.6, and a trimmed lognormal

curve of height  $5 \times \sigma$  (~450 cases). The peak of the lognormal curve is typically on the fifth or sixth day. We inject a mixture of the two outbreak signatures to illustrate the robustness of the algorithm combination. We repeat this test setting 20 times.

When combining control charts, the desired false alert rate is varied in the range  $\alpha \in \{0.01, 0.05, 0.1, 0.2\}$ . We set the threshold of the monitoring charts to meet the desired overall false alert rate  $\alpha$ , using 1 year of training data (referred to as the *experimental threshold*).

## 5.2 Results

### 5.2.1 Residuals Combination

In this section we compare four preprocessing methods, two simple methods and two combinations. The simple methods are Holt-Winters' exponential smoothing and linear regression. The two combination methods are a combination of Holt-Winters and linear regression residuals and a day-of-week variant of this combination. We then monitor each of the residual series with a Shewhart control chart. Because the preprocessing methods use different lengths of training data, we use the remaining dataset to compute the threshold that gives us the  $\alpha = 0.05$  false alert rate. The resulting true alert rate is shown in Figure 8-8, which displays the true alert rate distribution

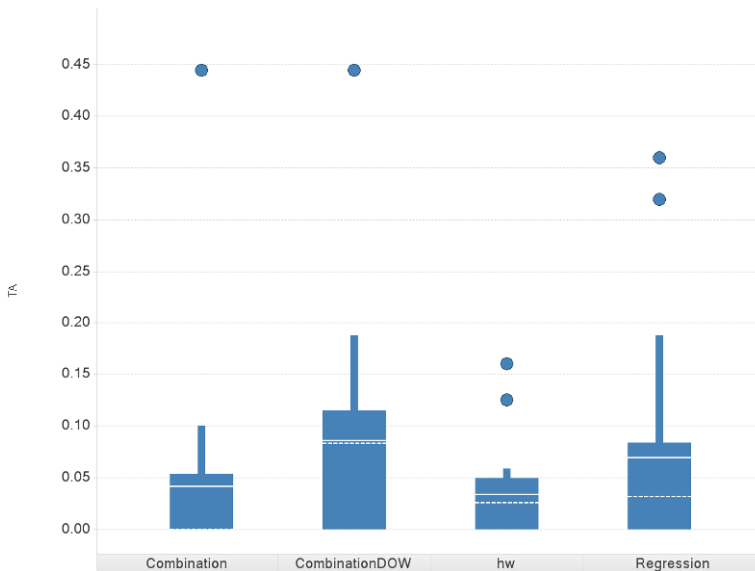


Figure 8-8. True alert rate distribution for Holt-Winters, linear regression, their combination, and a day-of-week combination. Means are marked by *solid white lines* and medians by *dashed white lines*.

for each of the four methods. Means and medians are marked as solid and dashed white lines, boxes correspond to the inter-quartile range, and the lines extend between the fifth and 95th percentiles.

The figure depicts the advantage of the day-of-week combined preprocessing, which has the highest mean, median and 75th percentile.

### 5.2.2 Control Chart Combination

We start by preprocessing the raw series using Holt-Winters exponential smoothing. Control charts (Shewhart, EWMA, CuSum, and the combined output) are then used to monitor the series of residuals. Finally, we calculate the false and true alert rates produced by each method. For the lognormal outbreak signature, we consider a true alert only if the alert took place before the peak, because timely alerting plays an important role in diminishing the spread of a disease.

In the first experiment we optimize the control chart combination separately for each of the 20 tests. Figure 8-9 depicts the results of this experiment. The different panels correspond to different levels of experimental threshold  $\alpha$ . Each panel shows the true alert rate distribution for each of the four methods. The results clearly show the advantage of the combined method in terms of both increasing true alert rate, subject to a given false alert rate, and in reducing the variance of the true alert rate.

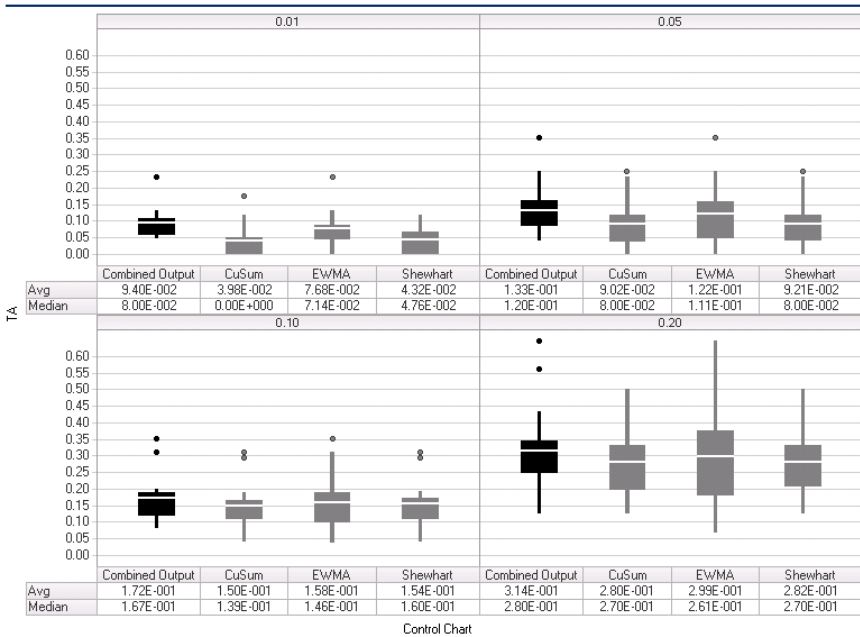


Figure 8-9. True alert rate distribution for three control charts and their combination, by false alert rate ( $\alpha = 0.01, 0.05, 0.10, 0.20$ ). Means are marked by solid white lines.

The main drawback of the first experiment is that the computation is very time consuming. Since achieving the optimal weights for the control charts is an NP complete problem, computation time increases exponentially in the length of the training data. Moreover, examining the actual weights shows that EWMA and Shewhart charts dominate the combination such that alerts are mostly determined by one of them (e.g., Shewhart) combined with an alert by one other method (e.g., either EWMA or CuSum). In an effort to reduce computation time, yet seek for good combinations, we take a hybrid approach: we choose among a small set of pre-determined combinations that appear to work well. This approach greatly reduces computation time and allows for real-time computation in actual settings.

Based on the general results found in the first experiment for the optimal weights, in the next experiment we chose two settings of pre-set weights:

1. Shewhart +: The algorithm signals an alert at time  $t$  if the Shewhart statistic signals an alert, *and* at least one other chart signals an alert.
2. EWMA +: The algorithm signals an alert at time  $t$  if the EWMA statistic signals an alert, *and* at least one other chart signals an alert.

The resulting true alert rates are shown in Figure 8-10. We observe that for a very low experimental false alert rate threshold ( $\alpha = 0.01$ ) the two new combination charts (Shewhart + and EWMA +) do not perform as well as the individual Shewhart and EWMA charts. However, when the experimental

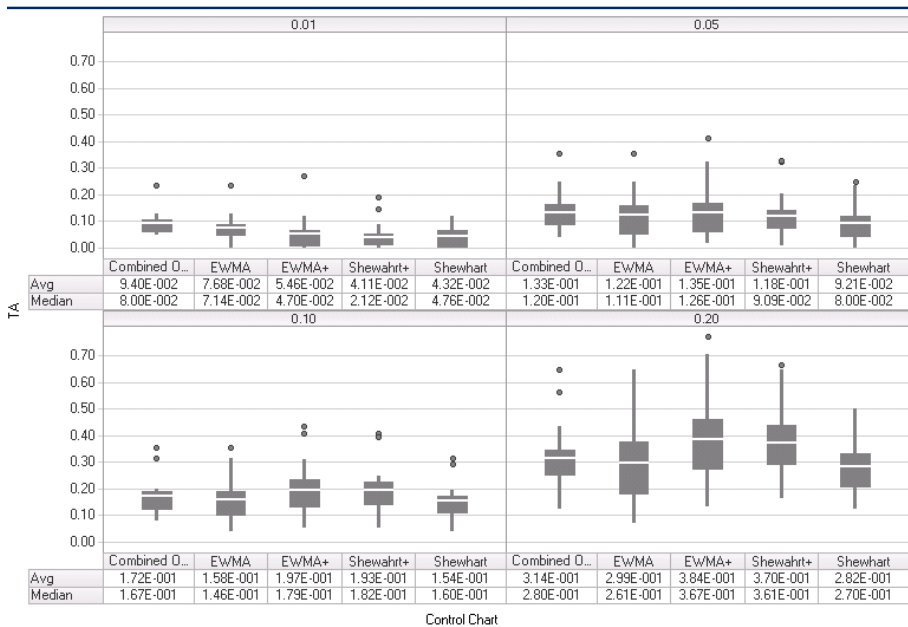


Figure 8-10. True alert rate distribution for select combinations, by false alert rate ( $\alpha = 0.01, 0.05, 0.10, 0.20$ ).

false alert rate threshold is higher ( $\alpha = 0.05$ ) the new charts perform at least as well as the ordinary charts, and even outperform the optimal combination (based on training data) when  $\alpha > 0.05$ . None of the methods violated the experimental false alert rate threshold by more than 10% when  $\alpha = 0.01$ , and 3% when  $\alpha \geq 0.05$ .

### 5.2.3 Combining Residuals and Monitoring

After examining the combination of residuals separately from combining control chart outputs, we now examine the effect of using combined preprocessing methods monitored by combined control charts on detection performance. The false alert rate is set to  $\alpha = 0.05$  and we observe the resulting true alert rate. We compare the performance of the different preprocessing methods monitored by either Shewhart, or Shewhart+. Figure 8-11 presents the resulting true alert rate distributions for each of the different combinations. We see that using Shewhart+ increases the true alert rate by approximately 50% compared to Shewhart. Also, the day-of-week residual combination outperforms the alternative preprocessing methods. The best performance is obtained by using both the day-of-week residual combination and applying the Shewhart+ to monitor the series. Thus, method combination provides improved detection at each of the preprocessing and monitoring levels, as well as when they are combined.

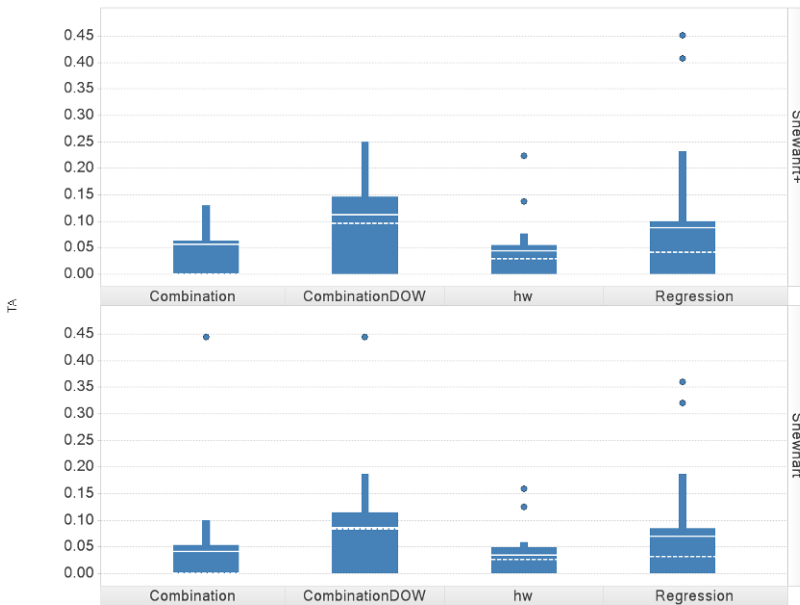


Figure 8-11. True alert rate distribution when combining residuals and control chart outputs (top panel is Shewhart+, bottom panel is Shewhart).

## 6. CONCLUSIONS

In this chapter we propose methods for improving the detection performance of univariate monitoring of non-stationary pre-diagnostic data by combining operations at each of the two stages of the outbreak detection task, data preprocessing and residual monitoring, for the purpose of increasing true alert rate for a given false alert rate.

Improved performance by combining control chart output is achieved by formulating the true alert/false alert tradeoff as a mixed integer programming problem (MIP). The MIP enables us to find the weights that optimize the combination method. To decrease computation time we take a hybrid approach where the weight optimization is carried out over a restricted set of combinations, which is obtained from a training stage. We show that the hybrid approach still provides improved performance. Our empirical experiments confirm the advantage of this portfolio approach in each of the stages (preprocessing and monitoring) separately and in the mixture of both.

Future extensions include adaptive combinations, where the weights of each method change dynamically over time, based on more current *history*. Another extension is using machine learning methods that automatically adjust combination weights based on current and recent *performance*, and on the most recent weight vector.

## ACKNOWLEDGEMENTS

The research was partially funded by NIH grant RFA-PH-05-126. Permission to use the data was obtained through data use agreement #189 from TRICARE Management Activity. For the second author, this research was performed under an appointment to the U.S. Department of Homeland Security (DHS) Scholarship and Fellowship Program, administered by the Oak Ridge Institute for Science and Education (ORISE) through an inter-agency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-AC05-06OR23100. All opinions expressed in this chapter are the authors and do not necessarily reflect the policies and views of DHS, DOE, or ORAU/ORISE.

## QUESTIONS FOR DISCUSSION

1. What is the role of data preprocessing in modern biosurveillance data? What are its limitations? How does preprocessing affect outbreak signatures in the data?



2. What are the requirements for combination methods?
3. How can the notion of combinations be extended for monitoring multiple time series? When would it be beneficial?
4. Discuss advantages and disadvantages of adaptive combinations, where the weights change over time.
5. What are other applications in which combination methods can be beneficial?

## REFERENCES

- Box, G., Luceno, A. (1997) *Statistical Control: By Monitoring and Feedback Adjustment*. 1st ed. Wiley-Interscience, London.
- Burkom, H. S. (2003) "Development, adaptation and assessment of alerting algorithms for biosurveillance", *Johns Hopkins APL Technical Digest* 24(4), 335–342.
- Burkom, H. S., Murphy, S. P., and Shmueli, G. (2007) "Automated time series forecasting for biosurveillance", *Statistics in Medicine* 26(22), 4202–4218.
- Lotze, T. and Shmueli, G. (2008) "Ensemble forecasting for disease outbreak detection", *23rd AAAI Conference on Artificial Intelligence, Chicago July 08*.
- Lotze, T., Shmueli G. and Yahav, I. (2007) Simulating Multivariate Syndromic Time Series and Outbreak Signatures, *Robert H. Smith School Research Paper No. RHS-06-054*, Available at SSRN: <http://www.ssrn.com/abstract=990020>.
- Lotze, T., Murphy, S. P., and Shmueli, G. (2008) "Preparing biosurveillance data for classic monitoring", *Advances in Disease Surveillance* 6. <http://www.isdsjournal.org/article/view/516>
- Montgomery, D.C. (1997) *Introduction to Statistical Quality Control*. 3rd ed. Wiley, London.
- NIST/SEMATECH: E-handbook of statistical methods. <http://www.itl.nist.gov/div898/handbook>.
- Siegrist, D. and Pavlin, J. (2004) Bio-ALIRT biosurveillance detection algorithm evaluation. *MMWR*, 53(Suppl), 152–158.
- Shmueli, G. and Burkom, H. S. (2010) "Statistical Challenges Facing Early Outbreak Detection in Biosurveillance", *Technometrics* vol 52(1), 39–51.

## ONLINE RESOURCES

<http://www.projectmimic.com>: R code for simulating or mimicking multivariate time series of biosurveillance data and outbreak signatures. Also includes semi-authentic data.