

# Directionally-Sensitive Multivariate Control Charts in Practice: Application to Biosurveillance

**Inbal Yahav**

Department of Information Systems  
School of Business Administration  
Bar Ilan University, Israel  
*inbal.yahav@biu.ac.il*

**Galit Shmueli**

Indian School of Business  
Gachibowli, Hyderabad 500 032, India  
*Galit\_Shmueli@isb.edu*

November 2, 2012

## **Abstract**

Multivariate control charts are used for monitoring multiple series simultaneously, for the purpose of detecting shifts in the mean vector in any direction. In the context of disease outbreak detection, interest is in detecting only an increase in the process means. Two practical approaches for deriving directional Hotelling charts are Follmann's correction and Testik and Runger's quadratic programming. However, there has not been an extensive comparison of their practical performance. Moreover, in practice many of the underlying method assumptions are often violated and the theoretically-guaranteed performance might not hold. In this work we compare the two directionally-sensitive approaches: a statistically-based approach and an operations research solution. We evaluate Hotelling charts as well as two extensions to multivariate EWMA charts. We examine practical performance aspects such as robustness to often impractical assumptions, the amount of data required for proper performance, and computational aspects. We perform a large simulation study and examine performance on authentic biosurveillance data.

keywords: Disease outbreak detection; Hotelling; Multiple Testing; Multivariate EWMA; Sensitivity Analysis

# Introduction and Motivation

## Motivating Application: Biosurveillance

Modern biosurveillance involves monitoring of a wide-range of pre-diagnostic and diagnostic data for the purpose of enhancing the ability of the public health infrastructure to detect, investigate, and respond to disease outbreaks. Statistical control charts have been a central tool in classic disease surveillance and have also migrated into modern biosurveillance. A major feature of biosurveillance data is multiplicity in several dimensions. The multiplicity of data sources (e.g., over-the-counter medication sales, nurse hotlines, and emergency department visits), multiple locations (e.g., multiple hospitals in a certain region), a variety of diseases of interest, multiple time series from a single source (e.g., medications for treating different symptoms), etc. In this work we focus on multiple time series arriving from a single data source, or from multiple data sources. A central question that arises is whether to monitor each series separately and then to combine the results in some fashion, or instead to monitor the series in a multivariate fashion. Current temporal monitoring in biosurveillance systems is done univariately, by applying univariate control charts to each series separately. This multiple testing can result in a very high false alert rate, leading many users to ignore alerts altogether. An alternative is to use multivariate control charts, which have traditionally been used in industry for monitoring multiple series simultaneously. This alternative to employ multiple univariate charts simultaneously helps avoid the multiple testing phenomenon. Furthermore, multivariate control charts take advantage of the correlation structure between individual series, thereby having a higher potential of detecting small signals that are dispersed across series. However, several theoretical and practical issues arise regarding the usefulness of multivariate control charts in biosurveillance. In particular, the characteristics of biosurveillance data and the conditions under which monitoring is performed usually mean that standard assumptions are not met, thereby rendering theoretical derivations questionable. This paper tackles the challenge of directional-sensitive monitoring *in practice*, in terms of the sensitivity and robustness of several multivariate monitoring methods for detecting outbreak signatures in multivariate biosurveillance-type data.

Fricker et al and Runger and Testik<sup>1,2,3</sup> offer comparisons between different implementations of multivariate CUSUM charts, Hotelling and multivariate CUSUM charts, and between multivariate CUSUM and multivariate EWMA (MEWMA) charts. They evaluate performance on simulated multivariate normal data with a seasonal sinusoidal cycle and a random fluctuation, as well as on authentic data. Our paper differs from the two papers above in that we examine the robustness of such methods to various assumption violations as well as to practical data conditions.

The novelty of this work is by isolating the properties of authentic data streams that violate the basic assumptions of the multivariate monitoring methods, and evaluating the sensitivity and robustness of the methods to these violations. Our paper focuses on the evaluation of computationally-feasible, model-based methods for monitoring multivariate data in a directionally-sensitive way. The main contribution is the evaluation of these methods in light of actual data characteristics and application conditions in disease outbreak detection, which often violate the assumptions of these methods. In addition, we assess the effects of practical factors such as the number of series, the amount of training data, the level of

cross-correlation, and data quality issues on performance.

This paper aims to fill the gap between methodology and practice, by assessing the feasibility and usefulness of methodologically-based tools for application in biosurveillance. Closing the gap between methodological research and practical application is extremely important in biosurveillance, where currently implemented systems are over-simplistic and where there is a reluctance to use “overly complicated” methods by public health users. One facet of this issue is the univariate monitoring used, and the familiarity with standard univariate control charts. In light of this environment, we chose to investigate the use of two relatively straightforward multivariate, directionally-sensitive extensions of known univariate charts (namely, Shewhart charts). We also extended them to MEWMA charts in the process, because those are also practically acceptable, and methodologically more appropriate in our setting. Our focus is on the conditions under which these different methods will be applied, and their practical limitations.

## Directional Monitoring

Three popular univariate control charts are the Shewhart chart, the Cumulative Sum (CuSum) chart, and the Exponentially-Weighted Moving Average (EWMA) chart. Multivariate extensions exist for each of these: The Hotelling  $T^2$  chart, the MCuSum and the MEWMA, respectively (see e.g. application to biosurveillance<sup>4,5,6,7,8,9</sup>). These multivariate charts are aimed at detecting a change in one or more of the process means in any direction. However, in many applications the interest is in detecting only an increase (or decrease) in one or more of the means. This is the case in biosurveillance, where series of daily diagnostic and pre-diagnostic counts are monitored for detecting a disease outbreak. The hypothesis is that an epidemic will manifest in the series as an increase in daily counts. In the univariate case there are simple corrections of the bi-directional charts to accommodate a one-directional change. In the multivariate case correcting for directional sensitivity is more complicated. One approach has been to empirically adjust the threshold of ordinary multivariate charts to achieve a given false alarm rate and then to evaluate its true alerting properties.<sup>2</sup> In this paper, we focus on two approaches that are useful for practical implementation: Follman<sup>10</sup> provides a correction for the ordinary Hotelling chart<sup>6</sup> and Testik and Runger<sup>11</sup> present a quadratic-programming approach to estimate the in-control mean vector. These two methods take two different approaches to yield directionally-sensitive Hotelling charts: The former is a statistical approach, while the latter is an operations-research approach. We describe each of these methods in detail and generalize them to obtain directionally-sensitive MEWMA charts. Using a large array of simulated data, we compare the performance of the directionally-sensitive Hotelling and MEWMA charts as a function of the number of monitored series, the cross-correlation structure, and the amount of training data required for estimating the covariance matrix. We then evaluate the robustness of the charts to underlying assumptions of normality and independence. Lastly, the four charts are applied to a set of authentic biosurveillance data. We summarize with conclusions and future research directions.

# Multivariate Control Charts

We use the following notation throughout the paper. Let  $\underline{X}_t = \{X_t^1, \dots, X_t^p\}$  be a  $p$ -dimensional multivariate normal vector with mean  $\underline{\mu} = \{\mu^1, \dots, \mu^p\}$  and covariance matrix  $\Sigma$ . We assume that at every time point  $t$ , a *single* observation is drawn from each of  $p$  series.

In the next sections, we describe the Hotelling control chart, followed by modifications by Follmann<sup>10</sup> and by Testik and Runger<sup>11</sup> for directional-sensitivity. We then expand their methods to obtain directionally-sensitive MEWMA charts.

## Hotelling's $T^2$ Control Chart

The multivariate extension of the ordinary Shewhart chart is the  $\chi^2$  chart, where the monitoring statistic is<sup>6</sup>

$$\chi_t^2 = (\underline{X}_t - \underline{\mu})' \Sigma^{-1} (\underline{X}_t - \underline{\mu}). \quad (1)$$

This is the squared statistical distance (also known as *Mahalanobis distance*) of the observation on day  $t$  from the in-control mean vector. Under the null no-shift hypothesis the statistic follows a  $\chi^2(p)$  distribution. The alarm threshold is therefore  $\chi_\alpha^2(p)$ , where  $\chi_\alpha^2(p)$  is the  $\alpha$  quantile of the  $\chi^2(p)$  distribution.

When  $\Sigma$  is unknown, it is estimated from data of length  $tr$  (referred to as *training data*). The estimated covariance matrix is denoted by  $S$ . The statistic, known as the Hotelling  $T^2$  statistic, is given by<sup>12</sup>

$$T_t^2 = (\underline{X}_t - \underline{\mu})' S^{-1} (\underline{X}_t - \underline{\mu}). \quad (2)$$

Under the null hypothesis of no shift,  $T^2 \sim \frac{p(tr+1)(tr-1)}{tr(tr-p)} F(p, tr-p)$ . The alerting threshold for the Hotelling  $T^2$  statistic is therefore

$$\frac{p(tr+1)(tr-1)}{tr(tr-p)} F_\alpha(p, tr-p), \quad (3)$$

where  $F_\alpha(p, tr-p)$  is the  $\alpha$  quantile from the  $F(p, tr-p)$  distribution.

## Multivariate EWMA

The standard univariate EWMA chart is based on the statistic

$$Z_t = \lambda X_t + (1 - \lambda) Z_{t-1}, \quad (4)$$

where  $0 < \lambda \leq 1$  is the smoothing parameter (typically chosen in the range  $[0.1, 0.3]$ ). Under the null hypothesis of no shift, this statistic follows a normal distribution with mean  $\mu$  and asymptotic variance  $s_{EWMA}^s = (\lambda/(2-\lambda))s^2$ , where  $s$  is the standard deviation of  $X_t$  estimated from historical data. The alerting thresholds are therefore  $\mu \pm k \times s_{EWMA}^s$ , where the constant  $k$  is commonly set to 3.

A multivariate extension of EWMA (MEWMA) is based on first creating an EWMA vector from each of the univariate  $p$  series<sup>7</sup>

$$\underline{Z}_t = \Lambda \underline{X}_t + (1 - \Lambda) \underline{Z}_{t-1}, \quad (5)$$

where  $\underline{Z}_t$  is the EWMA  $p$ -dimensional vector at time  $t$ ,  $\underline{X}_t$  is the  $p$ -dimensional observation vector, and  $\Lambda$  is a diagonal matrix with smoothing parameters  $\lambda_1, \dots, \lambda_p$  on the diagonal. The monitoring statistic is then

$$\underline{Y}_t = \underline{Z}_t' \Sigma_Z^{-1} \underline{Z}_t. \quad (6)$$

Lowry<sup>7</sup> showed that for a sufficiently large  $t$  (i.e., after a start up period) and for  $\lambda_1 = \dots = \lambda_p = \lambda$  the covariance matrix,  $\Sigma_Z$ , is given by

$$\Sigma_Z = \frac{\lambda}{2 - \lambda} \Sigma. \quad (7)$$

The alerting threshold is still  $\chi_\alpha^2(p)$ .

Note that for small  $t$ , the  $(k, l)$ th element in the covariance matrix  $\Sigma_Z$  is given by

$$Cov(Z_k, Z_l)_t = \lambda_k \lambda_l \frac{1 - (1 - \lambda_k)^t (1 - \lambda_l)^t}{\lambda_k + \lambda_l - \lambda_k \lambda_l} \sigma_{k,l}, \quad (8)$$

where  $\sigma_{k,l} = Cov(X_k, X_l)$ . If  $\lambda_1 = \dots = \lambda_p$ , then (8) reduces to

$$\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2t}) \Sigma. \quad (9)$$

## Directionally-Sensitive Multivariate Control Charts

In order to use multivariate control charts for detecting a parameter shift in *one* direction (an increase or decrease), there have been several approaches. One approach is to modify the non-directional multivariate monitoring statistic, as suggested by Follmann.<sup>10</sup> Another approach is to construct likelihood-ratio statistics for the alternative hypothesis. Most of the derivations along these lines are theoretical in nature and are hard to implement for more than 2 series. However, Testik and Runger,<sup>11</sup> taking an operations research approach, proposed an alternative formulation which can be implemented in practice for many series. We next describe these two approaches and extend them to obtain directionally-sensitive MEWMA charts.

### Follmann's Approach

Follmann<sup>10</sup> introduced a correction to the standard Hotelling statistic that adjusts it for directional sensitivity. The corrected statistic is given by

$$\chi_t^{2+} = (\underline{X}_t - \underline{\mu})' \Sigma^{-1} (\underline{X}_t - \underline{\mu}). \quad (10)$$

Note that  $\chi_t^{2+}$  in Follmann's notation is equivalent to the ordinary  $\chi_t^2$  statistic. The '+' sign indicates that we are interested in mean increases only. The vector  $\underline{X}_t$  in the equation is the sample mean vector at time  $t$ . An alert is triggered when  $\{\chi_t^{2+} > \chi_{2\alpha}^2(p)$  and  $\sum_{j=1}^p (X_t^j - \mu^j) > 0\}$ . This means that we require the sum of the elements of the mean vector at time  $t$  to exceed the series mean in order to alert.

Follmann proves that the procedure has type I error rate equal to  $2\alpha$  whether or not the covariance is known, and uses simulations to illustrate its power and to compare it to more complicated likelihood ratio tests.

Alber et al<sup>13</sup> show that the test statistic in (10) is not invariant to scale transformations. They propose a simple fix where instead of  $\Sigma$  the correlation matrix is used.

### Extending Follmann's Method to MEWMA charts

We extend the method proposed by Follmann in order to convert the ordinary MEWMA chart to a directionally-sensitive MEWMA chart. This is done by replacing  $\Sigma$  with  $\Sigma_Z$  in equation (10) (or equivalently, the original correlation matrix  $R$  with  $R_Z$ ), and  $\underline{X}_t$  by  $\underline{Z}_t$ . In other words, the alerting statistic is given by

$$\chi_t^2+ = \underline{Z}_t' \Sigma_Z^{-1} \underline{Z}_t, \quad (11)$$

and an alert is triggered when  $\{\chi_t^2+ > \chi_{2\alpha}^2(p) \text{ and } \sum_{j=1}^p (Z_t^j - \mu^j) > 0\}$ .

Additionally, if an alert is triggered on day  $\tau$ , we allow an option of restarting the control chart by setting  $\underline{X}_\tau = \underline{\mu}$ . The action of restarting reduces the “ringing effect” of the algorithm, by removing the sequence of alerts that follow an initial alert when there is a gradual increase in the vector mean.

A method similar to that of Follmann's was proposed by Joner et al<sup>14</sup> and Fricker et al.<sup>3</sup> The authors propose the following statistic, which is the maximum between the Lawry et al<sup>7</sup> MEWMA statistic and 0:

$$Z_t = \max\{\lambda(X_t - \mu_0) + (1 - \lambda)Z_{t-1}, 0\}. \quad (12)$$

The method is evaluated using simulated multivariate normal data, with simulated Poisson outbreaks. However, we do not include this method in the current sensitivity analysis.

### Testik and Runger's Quadratic Programming Approach

A different approach for obtaining directional sensitivity is based on deriving the monitoring statistic from the likelihood ratio, which is equal to the maximum likelihood under the alternative hypothesis of a directional shift (either positive or negative) divided by the null likelihood. Nüesch<sup>15</sup> showed that (twice the) log-likelihood ratio is

$$2l(\underline{\mu}) = \hat{\underline{\mu}} \Sigma^{-1} \hat{\underline{\mu}}, \quad (13)$$

where  $\hat{\underline{\mu}}$  is the maximum likelihood estimator. The monitoring statistic  $\chi_t^2 = \underline{X}_t' \Sigma^{-1} \underline{X}_t$  is therefore proportional to the likelihood ratio under the alternative hypothesis. Nüesch proposed  $l(\hat{\underline{\mu}})$  as an alternative monitoring statistic, where  $\hat{\underline{\mu}}$  maximizes the log likelihood. Testik and Runger<sup>11</sup> (abbreviated *TR*) showed that this can be formulated as an easily solvable quadratic programming problem if the data are first standardized

$$\widetilde{\underline{X}}_t = \Sigma^{-1/2} \underline{X}_t, \quad (14)$$

$$\widetilde{\underline{\mu}} = \Sigma^{-1/2} \underline{\mu}. \quad (15)$$

Nüesch also proved that the corresponding threshold is given by

$$P(\chi^2 > c^2) = \sum_{i=1}^p w(i)P(\chi_i^2 > c^2), \quad (16)$$

where  $w(i)$  is the probability that  $\hat{\underline{\mu}}$  has exactly  $i$  nonzero elements.  $\chi_i^2$  is a chi-squared random variable with  $i$  degrees of freedom, and  $c^2$  is a constant threshold value.

The problem has been to compute the weights  $\omega(1), \dots, \omega(p)$ . While theoretical derivations exist, they are typically hard to implement beyond  $p \geq 3$ . TR<sup>11</sup> obtained the weights empirically, by simulating  $p$ -dimensional multivariate normal data with  $\underline{\mu} = \hat{\underline{\mu}}$  and  $\Sigma$  and estimating the weights from the simulated data. It is important to note, however, that this approach assumes a known covariance matrix. The problem, in terms of  $\tilde{\underline{\mu}}$ , is then

$$\hat{\underline{\mu}}_t = \arg \min_{\tilde{\underline{\mu}} \geq 0} (\tilde{X}_t - \tilde{\underline{\mu}})' (\tilde{X}_t - \tilde{\underline{\mu}}). \quad (17)$$

### Extending TR's Method to MEWMA charts

We extend the method by TR<sup>11</sup> to obtain a directionally-sensitive MEWMA chart. This is achieved by replacing  $\Sigma$  with  $\Sigma_Z$  and  $\underline{X}_t$  by  $\underline{Z}_t$  in equations (14) and (15). In other words, our standardized data and means are

$$\tilde{Z}_t = \Sigma_Z^{-1/2} \underline{Z}_t, \quad (18)$$

$$\tilde{\underline{\mu}} = \Sigma_Z^{-1/2} \underline{\mu}, \quad (19)$$

and the problem in terms of  $\underline{\mu}_z$  is therefore

$$\hat{\underline{\mu}}_t = \arg \min_{\tilde{\underline{\mu}} \geq 0} (\tilde{Z}_t - \tilde{\underline{\mu}})' (\tilde{Z}_t - \tilde{\underline{\mu}}). \quad (20)$$

In this formulation we do not implement a restart condition for computational reasons. According to TR's approach the entire series is transformed prior to the monitoring action. A restart action would therefore require a re-transformation of the data after each alert, thereby increasing the run time by a factor equal to the number of alerts. An alternative on-the-fly implementation would apply the transformation on a daily basis (rather than in retrospect). This would not allow using the simple matrix operations for obtaining  $\hat{\underline{\mu}}_t$ , but it would easily incorporate the restart condition.

## Performance and Robustness Comparison

We set out to evaluate and compare the four different directionally-sensitive multivariate control charts: Hotelling and MEWMA using Follmann's method, and Hotelling and MEWMA using Testik and Runger's method. We compare their actual in-control performance as a function of the number of monitored series ( $p$ ), the covariance structure ( $\Sigma$ ) and their robustness to assumption violations that are likely to occur in practice. We first describe the simulation setup, and then examine the different factors and their effect on performance. Finally, we examine performance in the presence of a mean increase. We consider shifts of different magnitude, shape, and their presence in subsets of the series. We also discuss outbreak detection in the presence of mean decreases.

## Simulation Setup

We generate multivariate normal data and vary the level of correlation between series ( $\rho = 0.1, 0.3, \dots, 0.9$ ) and the number of dimensions ( $p = 2, 3, \dots, 20$ ). The length of each series is set to  $T = 1000$  time points. To compute a false alert rate, the number of false alerts is divided by  $T$ . The desired false alert rate is set to  $\alpha = 0.05$ . The threshold is computed according to equations 11 (Follmann) and 16 (TR). For each combination of  $\rho$  and  $p$ , 100 replications (i.e., 100 time series, each of length  $T = 1000$ ) are generated. This creates a distribution of false alert rates for each combination. We then examine the sensitivity and the robustness of the control charts as we change the simulation setting.

We use *R2.4.0* (<http://cran.r-project.org/>) to implement our simulation.

## Impact of cross-correlation and number of series

We start by evaluating the actual false alert rates of the different charts by assuming that  $\Sigma$  is known and given by

$$\begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & \dots & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

where  $\rho \in \{0.1, 0.2, \dots, 0.5\}$ . For the MEWMA charts we set  $\lambda_1 = \dots = \lambda_p = \lambda \in \{0.3, 0.5\}$ . Figures 1-2 compare the distribution of false alerts (FA) as a function of the number of series  $p$  (on the y-axes) and correlation  $\rho$  (across panels) for the four methods. In all cases the charts were set to an FA of  $\alpha = 0.05$ . The results are provided as side-by-side boxplots, with the mean FA represented as a solid dot; the whiskers extending to the 5% and 95% percentiles; and outliers represented as hollow dots. For all four methods, independently of  $\rho, p$  or  $\lambda$ , the FA rate has mean of 0.05 (as desired) with standard deviation of less than 0.01.

[Figure 1 about here.]

[Figure 2 about here.]

Overall we see that all charts produce false alerts that are centered around  $FA = 0.05$ . The inter-quartile range is approximately (0.04,0.06) for all methods, although the MEWMA charts have a slightly larger variance when  $\lambda$  is small ( $\lambda = 0.3$ ). Note that the number of series does not appear to affect the false alert rate. This is not surprising as the thresholds in all four methods are a function of  $p$ . Another interesting observation is that the distribution is very stable across the different correlation levels for all charts. Based on these results, in our next experiments we set  $\rho = 0.5$  and for MEWMA charts we set  $\lambda = 0.3$  (a popular choice in practice.)

## Robustness to assumptions

Next, we study the robustness of the four charts to assumption violations. We examine the length of training data required to estimate an unknown covariance matrix. We then relax



the assumption of normality of the underlying observations and examine the behavior of the charts when the series are autocorrelated and when the data come from a multivariate Poisson distribution (a common scenario with count data).

### Unknown covariance matrix

In this setting we vary the cross-correlation in the range  $\rho \in (0.1, 0.5, 0.9)$ . We assume that the covariance structure is unknown and is approximated from a training data ( $tr$ ) of varying length (using the Pearson method). We examine the false alert rate of the charts as a function of the length of the training data.

Unlike bidirectional monitoring methods, in which the F distribution is used when  $\Sigma$  is unknown,<sup>12</sup> the use of the F distribution in Follmann’s and TR’s methods is inadequate: In Follmann’s method, recall that the  $\chi^2$  statistic has a threshold of  $2\alpha$  rather than  $\alpha$  (see Equation 10). Therefore, replacing the  $\chi^2$  statistic with an F statistic requires an additional modification to the alerting threshold, which is not straightforward (see Equation 3). In TR’s method, the threshold is computed empirically. Theoretically, we could simply replace the  $\chi^2$  distribution in equation 16 with F distribution. However, we find that the  $\chi^2$  formulation results in a much lower false alert rate compared to the F distribution. We therefore use a  $\chi^2$  test in our simulation study.

Figures 3-5 compare the false alert rate of the four methods. The training data length varies from  $tr = 100$  time points (left panel) to  $tr = 500$  time plots (right panel). We see that Follmann’s Hotelling chart has a slightly lower average false alert rate than TR’s Hotelling. This improved performance is more significant when the training period is short (e.g.,  $tr < 300$ ) and the number of series is high, yet the variance is higher. Follmann’s MEWMA, on the other hand, has both a lower average false alarm rate and lower variance.

Overall, we observe that for  $p \leq 5$  the average false alert rate is centered around the desired threshold of  $FA = 0.05$ , independent of the training data length. This implies that 100 data points are sufficient to estimate the covariance matrix accurately. However, for  $p > 5$  the performance depends on the length of the training data. When  $tr = 100$  the false alert rate increases exponentially in the number of series. When  $tr = 200$  the rate increases linearly in the number of series. The average false alert rate converges to 0.05 only when the training data include at least 300 time points (almost one year of data, for daily series!).

Our results coincide with previous literature findings. Jensen et al<sup>16</sup> reviewed the effect of parameter estimation on control charts performance. They found that when using univariate EMWA charts, the smaller the value of  $\lambda$ , the larger the required sample size for ensuring performance similar to that of a chart based on known parameters. Jones et al<sup>17</sup> recommended using 100 samples of size  $n = 5$  for  $\lambda = 0.5$  and 400 samples of size  $n = 5$  for  $\lambda = 0.1$ . For Hotelling charts, Nedumaran and Pignatiello<sup>18</sup> recommended using sample sizes of at least 200 when the number of observations is  $n = 5$  and the dimension is  $p = 3$ .

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

## Autocorrelated series

In this setting we examine the impact of autocorrelation on the false alert rate of the different charts. It has been shown that biosurveillance daily time series tend to be autocorrelated.<sup>19,20</sup> While the literature includes methods for monitoring univariate, autocorrelated Poisson counts,<sup>21,22</sup> we are not aware of extensions to the multivariate case. We use the method in<sup>23</sup> to generate autocorrelated data with order 1 (AR(1)). We set the cross-correlation to  $\rho = 0.5$  and vary the autocorrelation coefficient in the range  $\theta \in \{0.05, 0.15, \dots, 0.35\}$ .

To understand the impact of autocorrelated series on chart performance, we first assume a known  $\Sigma$ . Figure 6 compares the four charts. The false alert rate of the Hotelling variants appears similar for both methods and centered around the desired threshold of  $FA = 0.05$ . In contrast, in the MEWMA charts the false alert rate increases significantly as  $\theta$  increases, and more so for TR's method. We observe, for example, that even when the number of series is  $p = 2$  and the cross-correlation is  $\theta = 0.35$ , Follmann's false alert rate is approximately 0.18 and TR's is 0.2. For  $p = 10$  series the corresponding rates are 0.4 and 0.65.

To reduce false alerts, we next examine a version of Follmann's MEWMA where the univariate EWMA statistics are restarted after alerts (i.e., if an alert is set on day  $t$ , we set  $Z_t = \mu$ ). As mentioned earlier, for computational reasons, we did not implement TR's method with restarting. Figure 7 shows that the false alert decreases by a factor of 4, compared to the MEWMA without restart (Figure 6, 3rd panel).

[Figure 6 about here.]

[Figure 7 about here.]

## Multivariate Poisson Data

We now relax the normal distribution assumption, as it is often violated in authentic data. Instead, we generate multivariate Poisson data (using the method by Yahav and Shmeuli<sup>24</sup>) with varying arrival rate  $\lambda \in \{1, 5, 10, 20\}$ . Biosurveillance data are typically daily count data. In some instances the counts are sufficiently high to justify normal-based control charts, while in other cases the counts might be too low. In low count situations a reasonable approximation that has been used in practice is a Poisson distribution.<sup>25,14</sup> Methods to handle such data for the univariate case is studied and analysed in several papers.<sup>26,27</sup> Examples are daily counts of cough complaints in a small hospital, or daily counts of school absences in a local high school. Using a multivariate Poisson structure enables us to evaluate the performance of the control charts in low-count data.

We evaluate and compare the methods when  $\Sigma$  is known (Figure 8) and observe that MEWMA outperforms Hotelling in terms of FA rate. The difference in performance is more pronounced as  $\lambda$  decreases, and when the number of series is large (roughly  $p \geq 5$ ). Both Follmann and TR methods have equivalent performance. Similar to the findings in Stoumbos and Sullivan,<sup>28</sup> we find that MEWMA charts are more robust to non-normality than Hotelling charts. As the authors suggest, the robustness property might later cause a decreased rate of true alerts for detecting spike outbreaks.

[Figure 8 about here.]

## Out of Control Performance

In this section we evaluate the performance of the four control charts in the presence of unexpected anomalies. Since the exact shape and magnitude of a disease outbreak manifestation in pre-diagnostic data is unknown, we consider two shapes that represent abrupt and incremental signatures of varying magnitudes. In particular, we consider one day spikes and multi-day lognormal increases. For each control chart we examine the rate of which it identifies true outbreaks (the ratio of detections, denoted TA), the time to detection (the number of time points until the first true alert), and the false alert rate.

### Injecting Outbreaks

We consider iid normal data with correlation  $\rho = 0.5$  and estimate the correlation matrix from a history of  $tr = 500$  time points that do not contain outbreak signatures. We vary the number of series in the set  $p \in \{4, 8, 12, 16\}$  and inject outbreak signatures into subsets of the  $p$  series of size  $s \in \{25\%, 50\%, 75\%, 100\%\} \times p$ .

**Spikes.** In the first experiment we inject spikes (i.e., single-day outbreak signatures) into the data. The magnitude of the spike varies in the range  $o \in \{0.5, 1, 1.5, \dots, 4\} \times \vec{\sigma}$ , where  $\vec{\sigma}$  is the series standard deviation. This means that the spike size is proportional to the series standard deviation (in our experiments we set  $\vec{\sigma} = 1$ ). In each experiment 20 spikes are injected into a subset of series at different time points, and the resulting true and false alert rates are computed.

Figures 9-10 show the true alert rate (TA) of the control charts when spikes of different magnitudes are injected into all  $p$  series. Figures 11-12 present the same for spikes injected into 25% of the series. For the MEWMA chart we evaluate the TA with and without restarting after an alert. As expected, the TA rate is higher for the Hotelling charts. Also, it appears that TR's method outperforms Follmann's in terms of true detections in both Hotelling and MEWMA charts. The difference in performance is more noticeable when the subset is small.

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

To further analyze the true alert rate and its determinants, we examine the relationship between true and false alerts while controlling for other factors (outbreak size  $o$ , subset size  $s$  and number of series  $p$ ). We use a linear regression model to explore the magnitude of the cross effect between false and true alert, as shown by eq. 21. While the relationship between the alert factors are not necessarily linear, the simplicity of this analysis enables us to clearly illustrate the increase of false alerts in the presence of true alerts.

$$TA = \beta_0 + \beta_1 \cdot FA + \beta_2 \cdot s + \beta_3 \cdot p + \beta_4 \cdot o + \epsilon. \quad (21)$$

Table 1 shows the output of the estimated model (p-values of the estimates are given in parentheses, coefficients significant at 5% are in bold). We find that controlling for all factors, TR’s Hotelling chart performs on average 30% better than Follmann’s in terms of percentage of true alert for every additional 1% in false alert rate. The performance is equal only when the subset size is close to 100%. Similar results are observed for MWEMA without restart. Another observation is the strong correlation between TA and FA in the MEWMA control charts. Controlling for all other factors, a 1% increase in FA rate results in an average 5.35% (Follmann) and 4.24% (TR) increase in TA. MEWMA with restart, on the other hand, has the exact opposite relationship, presumably since the restarting action erases the history and allows the control chart to re-accumulate small deviations from the means.

**Log-Normal Signatures.** Next, we inject into the data multi-day stochastic gradual increases taken from a lognormal distribution. As with the spikes, we vary the magnitude of the signature and the fraction of ‘infected’ series. We inject a single signature in each experiment and examine the number of time points until the first successful detection as well as the true alert rate. Results are shown for subset size  $s = 25\%p$  in Figures 13-14. Panels correspond to different magnitudes (we plot partial magnitudes for brevity). Boxplots show the distribution of time to detection (black dots are average time to detection), *conditional* on the outbreak being detected.

We observe that MEWMA charts perform better than Hotelling charts. This result is expected, as MEWMA charts are designed to detect gradually increasing signals. We again see that TR’s method outperforms Follmann’s both in terms of TA and time to detection.

[Figure 13 about here.]

[Figure 14 about here.]

## Detection of mean increases in the presence of mean decreases

Although we are interested in detecting only increases in the process mean, it is possible that due to data quality anomalies, one or more of the means will decrease (e.g., due to reduced reporting or problems with data recording on a certain day). We therefore evaluate the performance of the control charts when we inject both *positive* and *negative* spikes, where negative spikes represent such anomalies. Recall, however, that for detecting disease outbreaks rather than data quality anomalies, we are only interested in detecting mean *increases* (i.e., positive spikes).

We consider a bivariate dataset ( $p = 2$ ). In each experiment we inject 20 *positive* spikes into the first series and 20 *negative* spikes into the second series. *Positive and negative spikes are injected on the same days*. The magnitude of the spikes varies in the range  $o \in \{1, 2, 3\} \times \vec{\sigma}$ , where  $\vec{\sigma}$  is the series standard deviation.

Figure 15 (top panels) depicts the resulting true alert detection rate of the four methods. For comparison, we repeat the same experiments with positive spikes only (middle panels) and with negative spikes only (injected into a single series, bottom panels). Note that ‘True Alert’ (TA) in the bottom panels refers to the detection rate of decreases in the mean. We see that, surprisingly, TR’s Hotelling and MEWMA charts detect mean decreases and alert

as if they were also outbreaks (i.e., mean increases). The result is an ‘improved’ performance when both positive and negative spikes are present, yet poor performance (increased false detection) when only negative spikes are injected into the data.

[Figure 15 about here.]

As for Follmann’s charts, a simultaneous decrease and increase in the mean vector leads to poor performance of Follmann’s Hotelling chart. This result is expected, as Follmann’s method alerts when the summation of the series is greater than zero (see Equation (11)). In contrast, although Follmann’s MEWMA chart performs similar to the Hotelling chart, it performs equally well in the presence and the absence of decreasing spikes.

## Results for Authentic Data

We now examine the behavior of the four control charts when applied to authentic bio-surveillance data. The data include series of daily counts of patients arriving at emergency departments in a certain US city, between Feb-28-1994 and Dec-30-1997, broken down by the type of “chief complaint”. The counts are grouped into 13 categories using the CDC’s (Centers for Disease Control and Prevention) syndrome groupings. The data are shown in Figure 16.

1

[Figure 16 about here.]

Authentic data such as these typically contain several explainable patterns like seasonality and day of week effects, which clearly violate the assumption of classic control charts. We therefore first preprocess the series to remove these patterns. There are different methods for performing this step.<sup>23,19</sup> Here we use Holt-Winter’s exponential smoothing to remove day-of-week effects, seasonality, and autocorrelation. We obtain a series of forecast errors<sup>19</sup> that better adhere to control chart assumptions. One of the main challenges with authentic pre-diagnostic data is that they are unlabeled, such that outbreak periods are usually unknown. For purposes of evaluation we therefore assume that the authentic data *do not contain any signatures of unusual diseases*, and the only signatures are those that we inject artificially. This assumption is reasonable when the goal is to detect disease outbreaks that we know are not present in the data (such as an outbreak following a bioterrorist attack or a pandemic such as avian flu or SARS). The assumption is not reasonable if the goal is to detect an event such as the onset of Influenza which recurs annually. In our case we are indeed interested in detecting unknown disease outbreaks and hence the assumption is reasonable.

We compute the FA rate by applying the four control charts to the preprocessed data. To evaluate TA rate and time to detection, we inject 32 spikes of magnitude  $o$ , where  $o \sim u[1, 4] \times \vec{\sigma}$  into a random subset of the 13 series. The covariance structure is estimated from the first year of data (365 time points). Applying each control chart to the data with

---

<sup>1</sup>We thank Dr. Howard Burkom of the Johns Hopkins University’s Applied Physics Laboratory, for making this aggregated dataset, previously authorized by ESSENCE data providers for public use at the 2005 Syndromic Surveillance Conference Workshop, available to us.

injected signatures we compute their FA and TA rates. The results are shown in Table 2. There are two main observations: (1) Follmann’s MEWMA chart with restart alerts the least, whether or not there are outbreak signatures, and (2) TR’s Hotelling chart is most sensitive: it has the highest TA rate, but also the second highest FA rate. The highest FA rate is obtained with TR’s MEWMA, but this is likely due to the lack of restart after an alert. Note also that the FA rate computed before and after the signature injections are similar.

To further explore these two results and the relationship between TR’s Hotelling and Follmann’s MEWMA (with restart), we examine their performance across a range of FA rates ( $[0, 0.2]$ ). For a higher sensitivity comparison, we examine only outbreaks of smaller magnitude ( $o \sim u[0.5, 2.5] \times \bar{\sigma}$ ). Results are shown in Figure 17. We see that the low FA rate is controlled by Follmann’s MEWMA chart and the high TA rate is controlled by TR’s Hotelling chart. These results are in line with those obtained from the simulated data. The conclusion is therefore that the choice of chart should be driven by the tradeoff between true and false alerts required by the user.

[Figure 17 about here.]

Finally, to evaluate the advantages of each of the four multivariate control charts we compare them against univariate monitoring where univariate Shewhart charts are applied simultaneously to each of the  $p = 13$  series. The rule for alerting is when at least one of the charts alerts. We vary the *actual* FA rate of the multiple-univariate charts between  $[0, 0.6]$  and observe the TA rate. Results are shown in Figure 18. We can see that all of the multivariate charts are Pareto efficient compared to the multiple-univariate Shewhart.

[Figure 18 about here.]

## Conclusions and Future Directions

We present and evaluate four tools for monitoring multivariate time series for the purpose of detecting anomalies that manifest in a certain direction. The directionally-sensitive multivariate control charts are Follmann’s Hotelling, TR’s Hotelling, Follmann’s MEWMA (with and without restart) and TR’s MEWMA. All charts have underlying assumptions such as normality, independence, and knowledge of the covariance structure, which rarely hold in practice. We therefore evaluate and compare their performance when each of the assumptions is violated. We also examine practical issues such as length of the training set, the number of monitored series, the effect of data quality issues such as mean decreases, and the subset of series in which the outbreak signature appears. All these are manipulated using simulation, where we can assess the impact of each factor separately. Finally, we apply the charts to authentic data (with and without injected outbreak signatures) and compare their TA and FA rates.

Note that control charts are often applied to raw data rather than pre-processed data in practice. Yet raw data usually violate the normality assumption and also exhibit high levels of autocorrelation. In that case the more robust method (Follmann’s) would be preferable. However, clearly the correct approach is to first pre-process the data.

The analysis in this paper is aimed at providing guidelines to biosurveillance systems where multiple time series are monitored. For a given dataset, and based on its characteristics (cross correlation, autocorrelation, etc.) and features (number of series, length of training data, etc.), we can evaluate the performance of each of the multivariate charts in terms of expected false and true alert rates and time to detection. These, in turn, can be used to choose one chart according to costs associated with missed and false alerts. To allow wide implementation of the tools and their incorporation in existing systems, and to be able to compare existing tools to the proposed multivariate charts, we make our code available online (<http://www.rhsmith.umd.edu/faculty/phd/inbal>).

There are several directions for extending this work. First, our simulated data are generated from a mean and covariance structure that do not change over time. In practice, however, data characteristics are subject to changes. To overcome this problem, one can consider estimating the mean and covariance structure repeatedly over time, using a moving window. Our framework is helpful in determining the length of this window.

Secondly, in terms of performance evaluation, once we move from the synthetic environment to authentic data, we no longer have replications of the series. This means that we cannot assess the statistical significance of the difference between the performance of the different charts (e.g., is TR's Hotelling's  $TA=0.88$  significantly higher than Follmann's MEWMA  $TA=0.84$ ?). Lotze et al<sup>23</sup> propose an elegant approach for creating multiple realizations of authentic data by 'mimicking' the statistical characteristics of an authentic multivariate set of data. Comparing performance over a sample of mimics would then enable assessing statistical significance of differences in false and true alerts as well as time to detection.

Finally, our focus was on comparing four multivariate monitoring tools, and we only briefly touched upon the comparison with multiple-univariate monitoring. Yahav and Shmueli<sup>29</sup> explore methods for combining univariate algorithms. A more thorough comparison is needed in order to assess under what conditions multivariate monitoring should be preferred over multiple univariate monitoring in practice.

## References

1. Runger GC, Testik MC. Multivariate extensions to cumulative sum control charts. *Quality and Reliability Engineering International* 2004; **20**(6):587-606.
2. Fricker RD. Directionally sensitive multivariate statistical process control procedures with application to syndromic surveillance. *Advances in Disease Surveillance* 2007; **3**:1-17.
3. Fricker RD, Knitt MC, Hu CX. Comparing directionally sensitive MCUSUM and MEWMA procedures with application to biosurveillance. *Quality Engineering* 2008; **20**(4):478-494.
4. Crosier RB. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* 1988; **30**(3):291-303.
5. Pignatiello JJ, Runger GC. Comparisons of multivariate CUSUM charts. *Journal of Quality Technology* 1990; **22**(3):173-186.
6. Hotelling H. *Techniques of Statistical Analysis*, chapter Multivariate Quality Control 1947; 113-184.
7. Lowry CA, Woodall WH, Champ CW, Rigdon SE. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* 1992; **34**(1):46-53.
8. Lowry CA, Montgomery DC. A review of multivariate control charts. *IIE Transactions* 1995; **27**(6):800-810.
9. Fricker R. Introduction to statistical methods for biosurveillance. *Cambridge University Press. Draft available online at August* 2010; **9**:869-8414.
10. Follmann D. A Simple Multivariate Test for One-Sided Alternatives. *Journal of the American Statistical Association* 1996; **91**(434):854-861.
11. Testik MC, Runger GC. Multivariate one-sided control charts. *IIE Transactions* 2006; **38**(8):635-645.
12. Montgomery DC, Klatt PJ. Economic Design of  $T^2$  Control Charts to Maintain Current Control of a Process. *Management Science* 1972; **19**(1):76-89.
13. Albers W, Kallenberg WCN, Martini F. Data-driven rank tests for classes of tail alternatives. *Journal of the American Statistical Association* 2001; **96**(454):685-696.
14. Joner MD, Woodall WH, Reynolds MR, Fricker RD. A one-sided MEWMA chart for health surveillance. *Quality and Reliability Engineering International* 2008; **24**(5):503-518.
15. Nüesch PE. On the Problem of Testing Location in Multivariate Populations for Restricted Alternatives. *The Annals of Mathematical Statistics* 1966; **37**(1):113-119.



16. Jensen WA, Jones-Farmer LA, Champ CW, Woodall WH. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology* 2006; **38**(4):349-364.
17. Jones IA, Champ CW, Rigdon SE. The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics* 2001; **43**(2):156-167.
18. Nedumaran G, Pignatiello JJ. On constructing  $T^2$  control charts for on-line process monitoring. *IIE transactions* 1999; **31**(6):529-536.
19. Burkom HS, Murphy SP, Shmueli G. Automated time series forecasting for biosurveillance. *Statistics in Medicine* 2007; **26**(22):4202-4218.
20. Shmueli G, Burkom HS. Statistical challenges facing early outbreak detection in bio-surveillance. *Technometrics* 2010; **52**(1):39-51.
21. Weiß CH. Controlling correlated processes of Poisson counts. *Quality and Reliability Engineering International* 2007; **23**(6):741-754.
22. Yontay P., Weiß CH., Testik MC., Pelin Bayindir Z. A Two-Sided Cumulative Sum Chart for First-Order Integer-Valued Autoregressive Processes of Poisson Counts. *Quality and Reliability Engineering International* 2012; Forthcoming.
23. Lotze T, Shmueli G, Yahav I. Simulating and evaluating biosurveillance datasets. *Bio-surveillance: A Health Protection Priority* 2010; 23-52.
24. Yahav I, Shmueli G. On generating multivariate poisson data in management science applications. *Applied Stochastic Models in Business and Industry*, 2011; online library.
25. Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology* 2004; **159**(3):217-224.
26. Han SW., Tsui KL., Ariyajunya B., Kim SB. A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. *Quality and Reliability Engineering International* 2009; **26**(3):279-289.
27. Sellers KF. A generalized statistical control chart for over-or under-dispersed data. *Quality and Reliability Engineering International* 2011; **28**(1):59-65.
28. Stoumbos ZG, Sullivan JH. Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology* 2002; **34**(3):260-276.
29. Yahav I, Shmueli G. Algorithm Combination for Improved Performance in Biosurveillance Systems. *Lecture Notes in Computer Science* 2007; **4506**:91-102.

[Table 1 about here.]

[Table 2 about here.]

## List of Figures

1	Distribution of false alert rate (FA) in directionally sensitive Hotelling charts as a function of the number of series $p$ and correlation $\rho$ . The charts were all set to FA=0.05 . . . . .	20
2	Distribution of false alert rate (FA) in directionally sensitive MEWMA as a function of the number of series $p$ and correlation $\rho$ . The charts were all set to FA=0.05 . . . . .	21
3	Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length (tr), when $\rho = 0.1$ . . . . .	22
4	Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length (tr), when $\rho = 0.5$ . . . . .	23
5	Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length (tr), when $\rho = 0.9$ . . . . .	24
6	Distribution of false alert rates (FA) in directionally-sensitive charts as a function of the autocorrelation ( $\theta$ ), when the covariance matrix is known . . . . .	25
7	Distribution of false alert rates (FA) in Follmann's directionally-sensitive MEWMA chart with restarts, as a function of the autocorrelation ( $\theta$ ), when the covariance matrix is known . . . . .	26
8	Distribution of false alert rates (FA) in directionally-sensitive charts for Poisson counts, as a function of the Poisson parameter ( $\lambda$ ), when the covariance matrix is known . . . . .	27
9	Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of spike magnitude . . . . .	28
10	Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of spike magnitude . . . . .	29
11	Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of spike magnitude when spike is injected into 25% of the series . . . . .	30
12	Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of spike magnitude when spike is injected into 25% of the series . . . . .	31
13	Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of outbreak magnitude when the outbreak is injected into 25% of the series . . . . .	32
14	Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of outbreak magnitude when the outbreak is injected into 25% of the series . . . . .	33
15	Distribution of true alert (TA) rate in directionally-sensitive charts, as a function of spike magnitude in the presence of increasing and decreasing spikes (top), increasing spikes only (middle) and decreasing spikes only (bottom) . . . . .	34
16	Authentic Data: Daily counts of chief complaints by patients arriving at emergency departments in a US city . . . . .	35
17	True vs. false alert rates for TR's Hotelling chart vs. Follmann's MEWMA chart with restarts . . . . .	36
18	True vs. false alert rates; comparing multivariate control charts with multiple-univariate Shewhart charts . . . . .	37

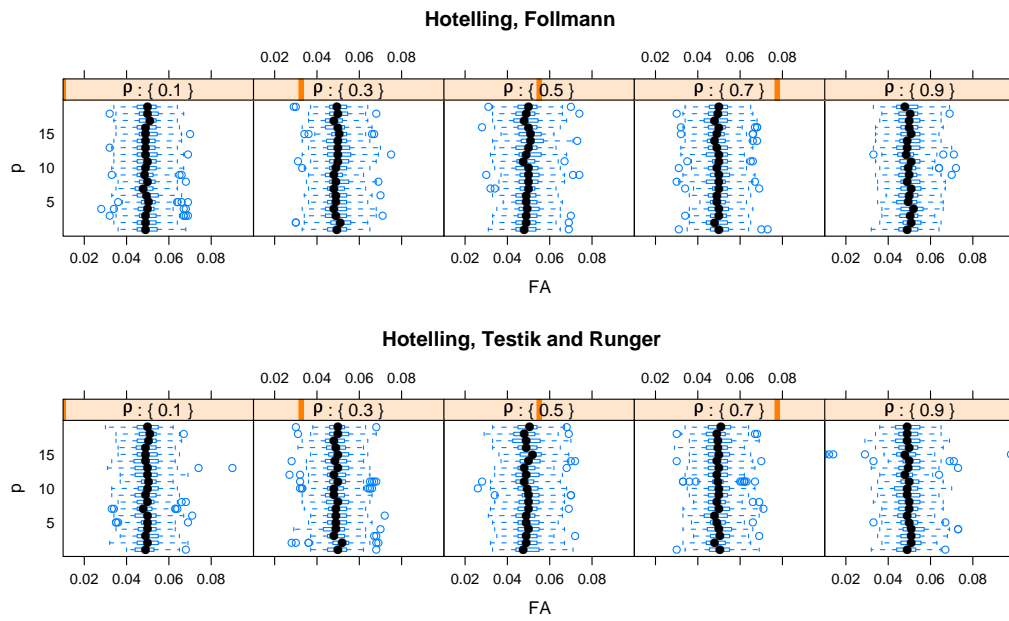


Figure 1: Distribution of false alert rate (FA) in directionally sensitive Hotelling charts as a function of the number of series  $p$  and correlation  $\rho$ . The charts were all set to FA=0.05

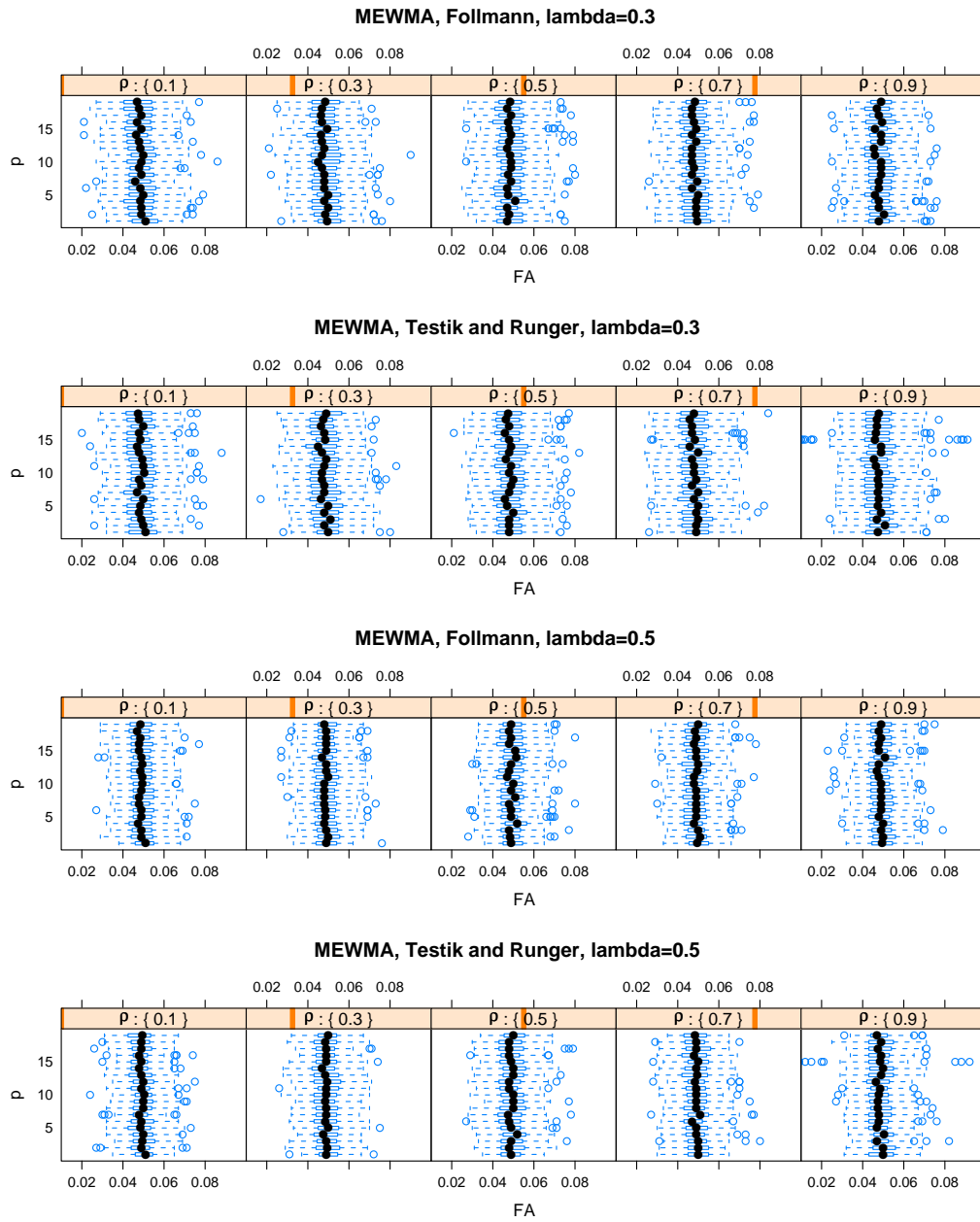


Figure 2: Distribution of false alert rate (FA) in directionally sensitive MEWMA as a function of the number of series  $p$  and correlation  $\rho$ . The charts were all set to FA=0.05

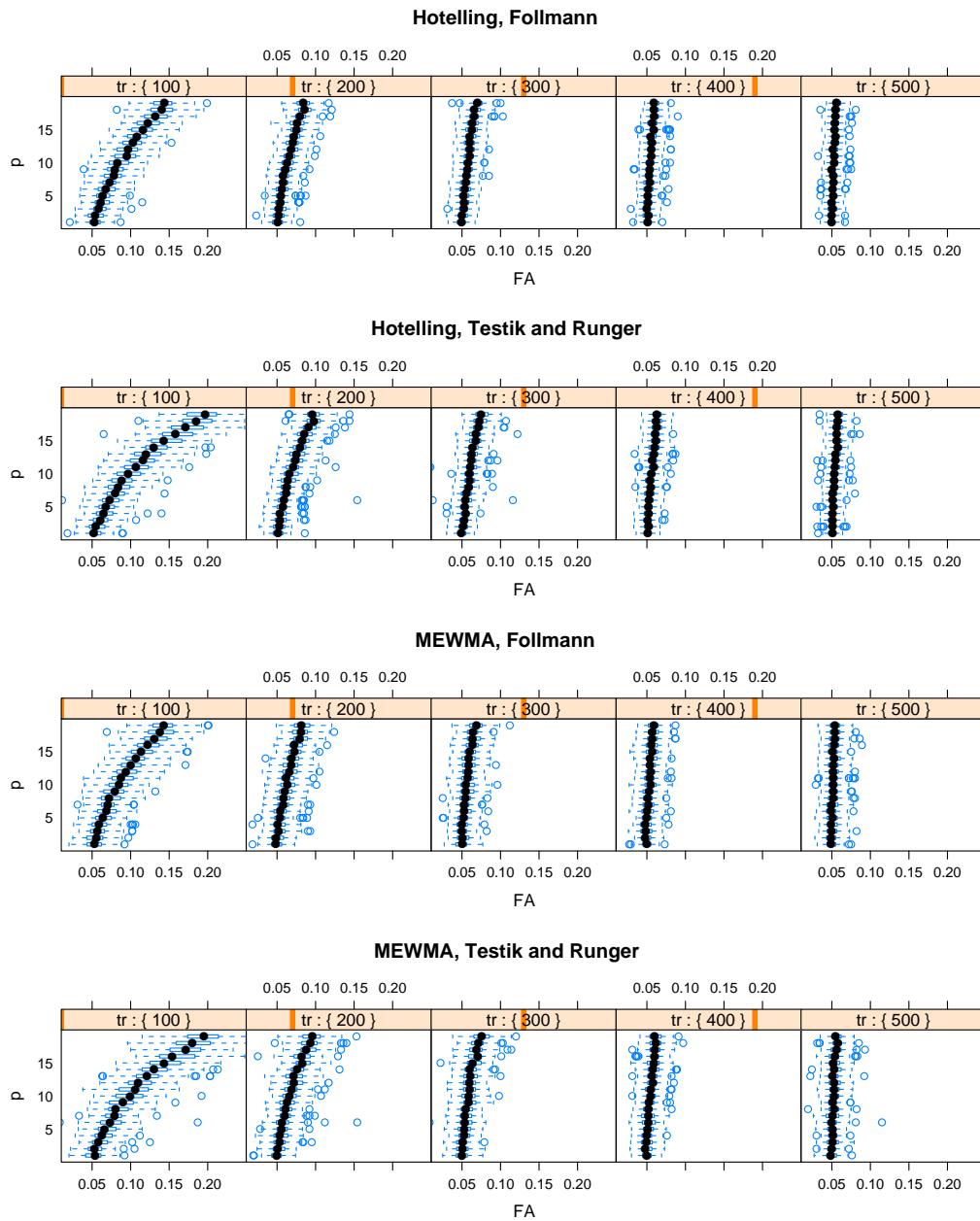


Figure 3: Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length ( $tr$ ), when  $\rho = 0.1$

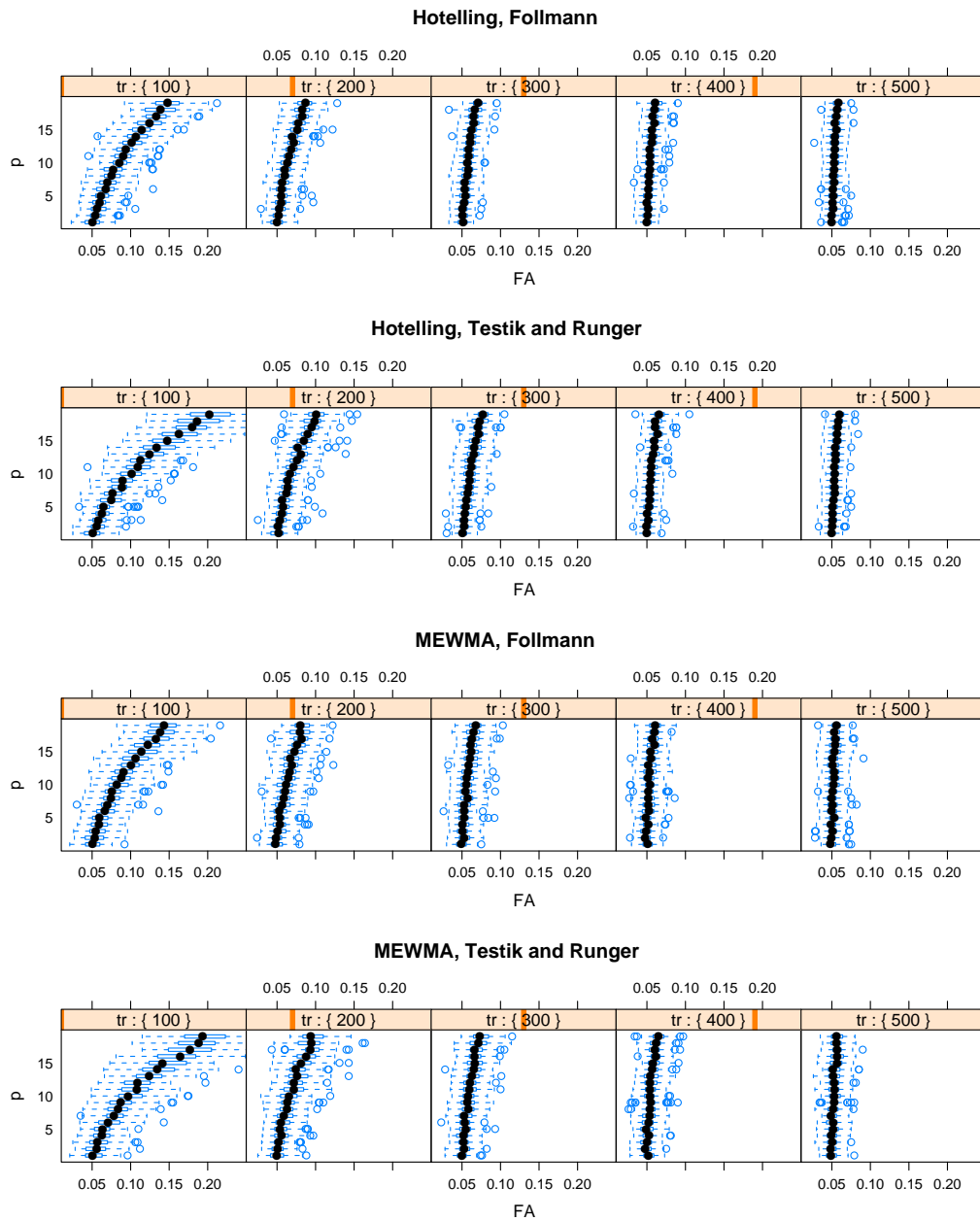


Figure 4: Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length (tr), when  $\rho = 0.5$

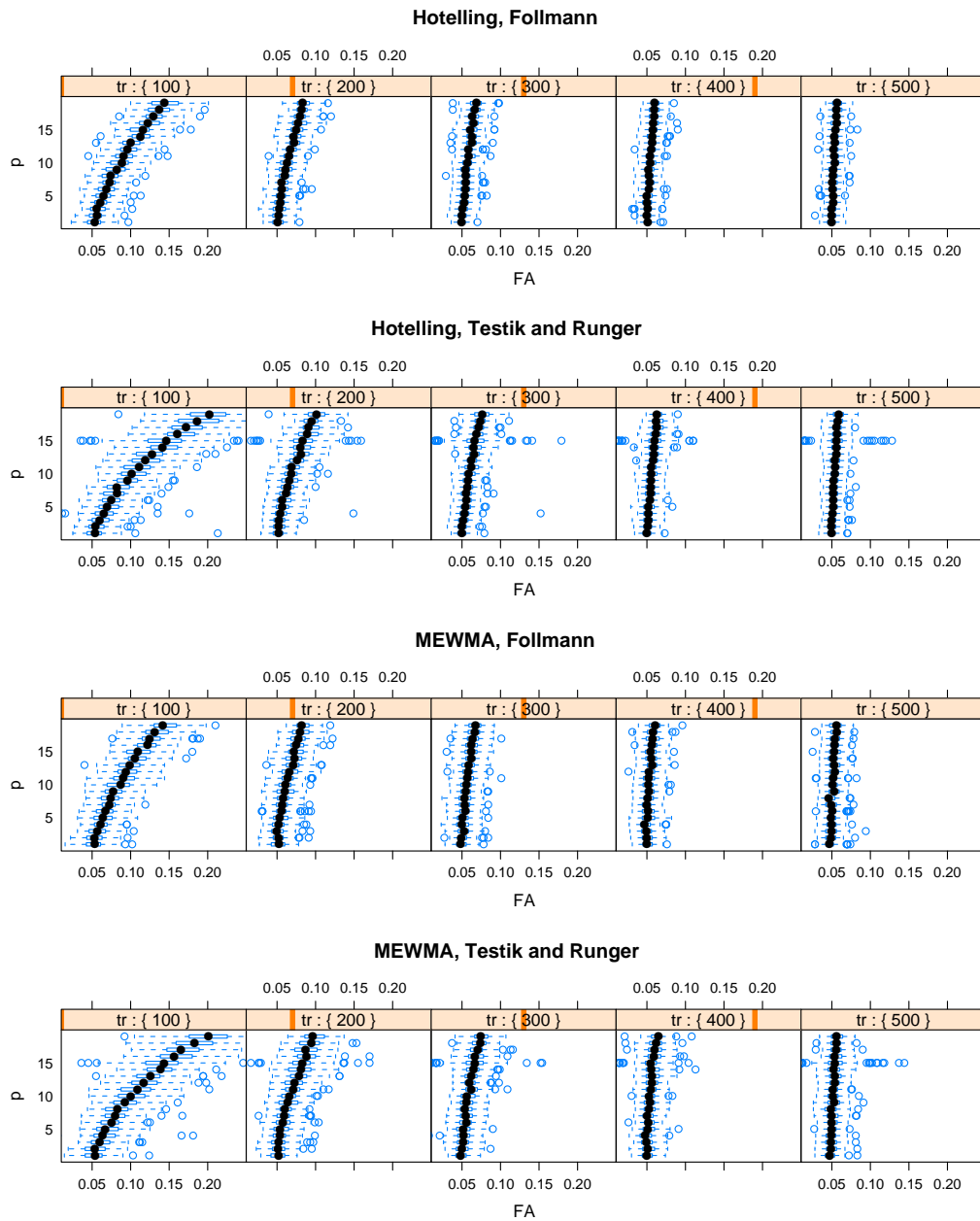


Figure 5: Distribution of false alert rate (FA) in directionally sensitive charts as a function of training data length ( $tr$ ), when  $\rho = 0.9$



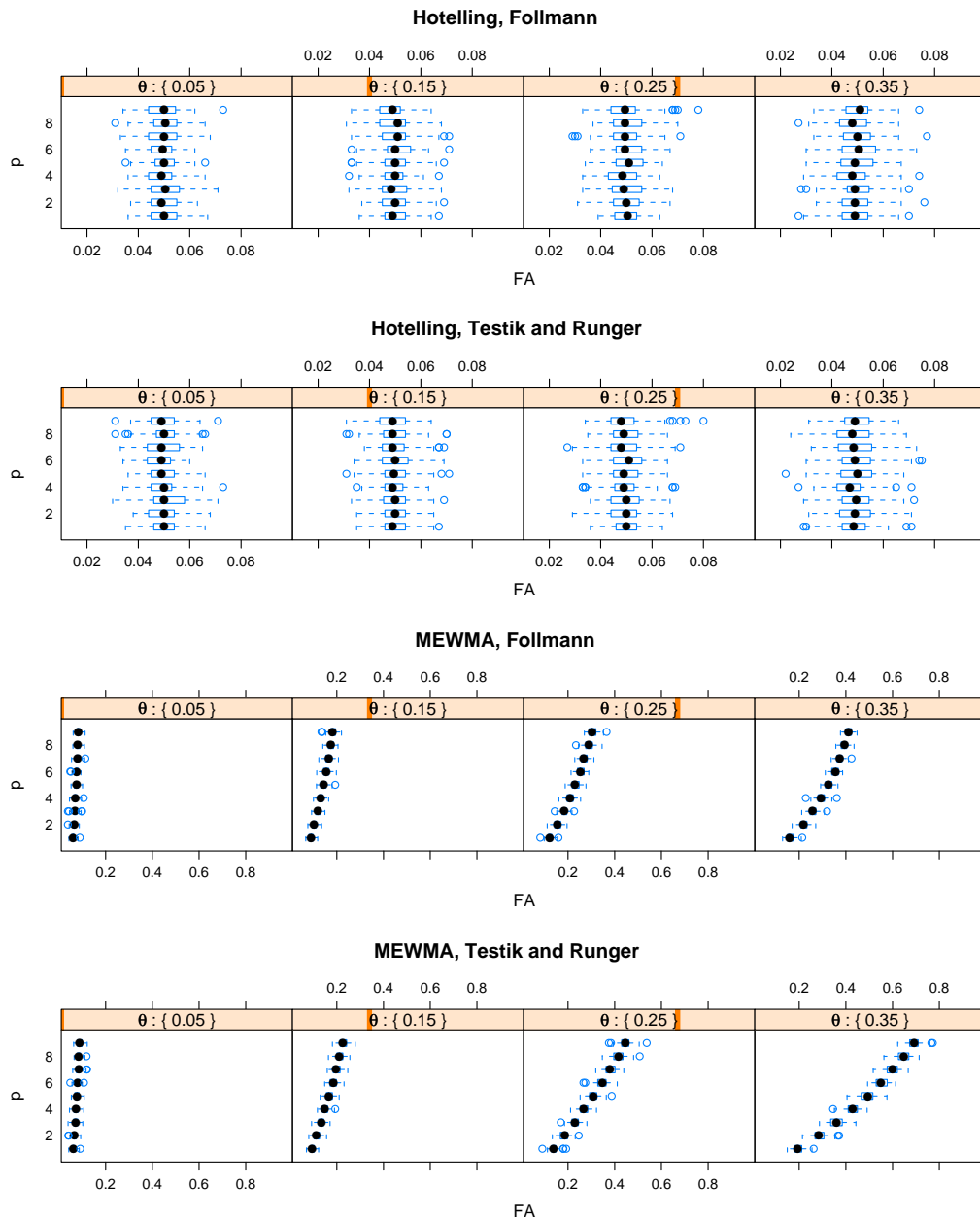


Figure 6: Distribution of false alert rates (FA) in directionally-sensitive charts as a function of the autocorrelation ( $\theta$ ), when the covariance matrix is known

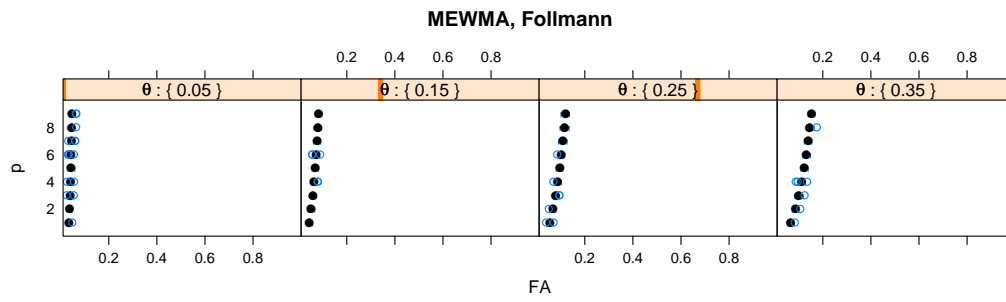


Figure 7: Distribution of false alert rates (FA) in Follmann's directionally-sensitive MEWMA chart with restarts, as a function of the autocorrelation ( $\theta$ ), when the covariance matrix is known

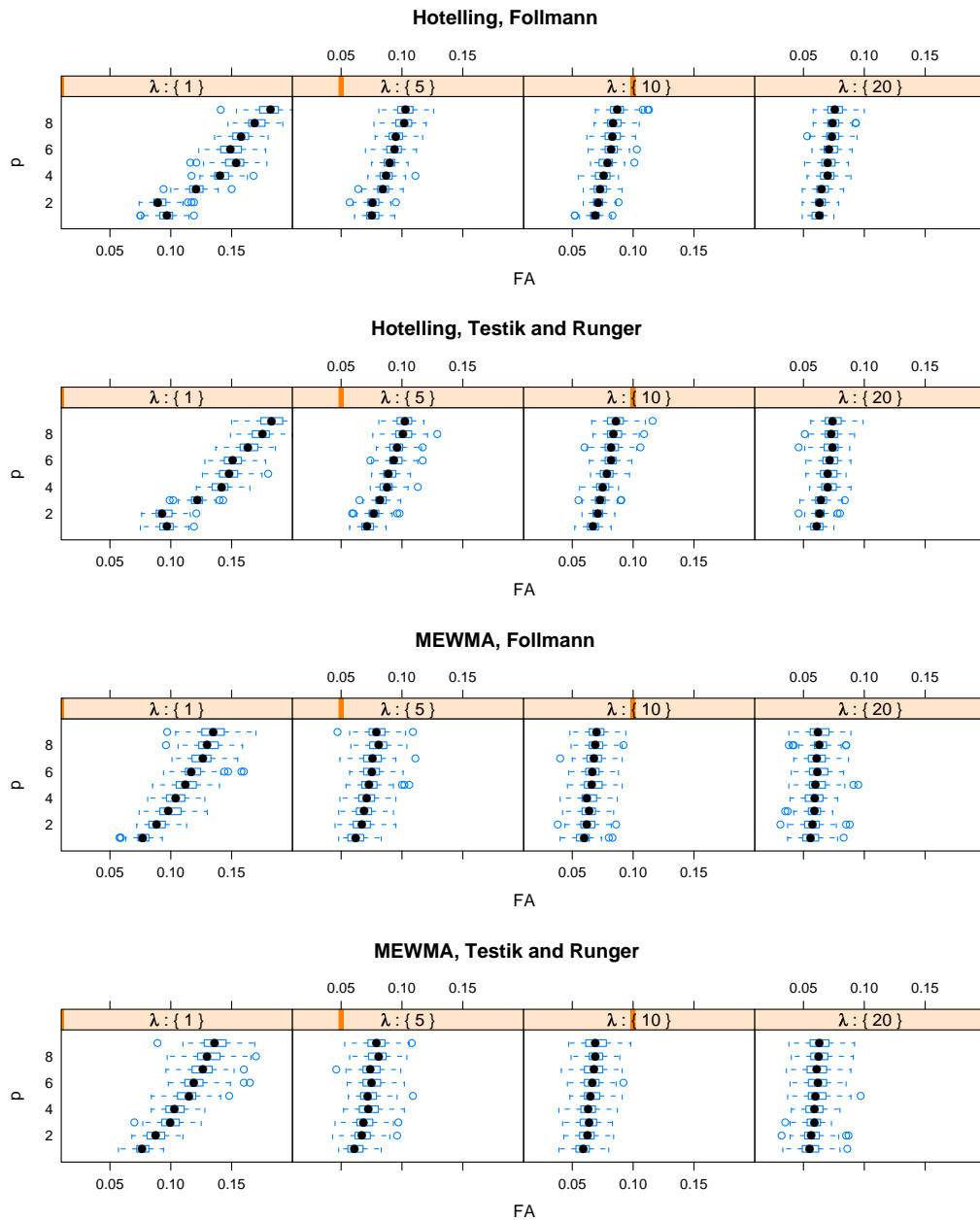


Figure 8: Distribution of false alert rates (FA) in directionally-sensitive charts for Poisson counts, as a function of the Poisson parameter ( $\lambda$ ), when the covariance matrix is known

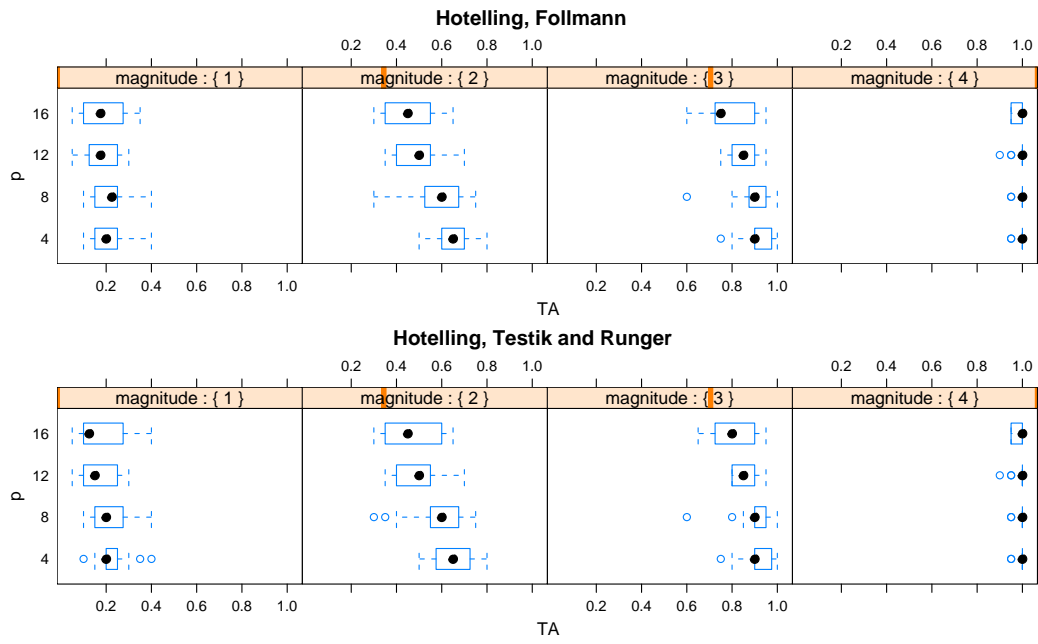


Figure 9: Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of spike magnitude

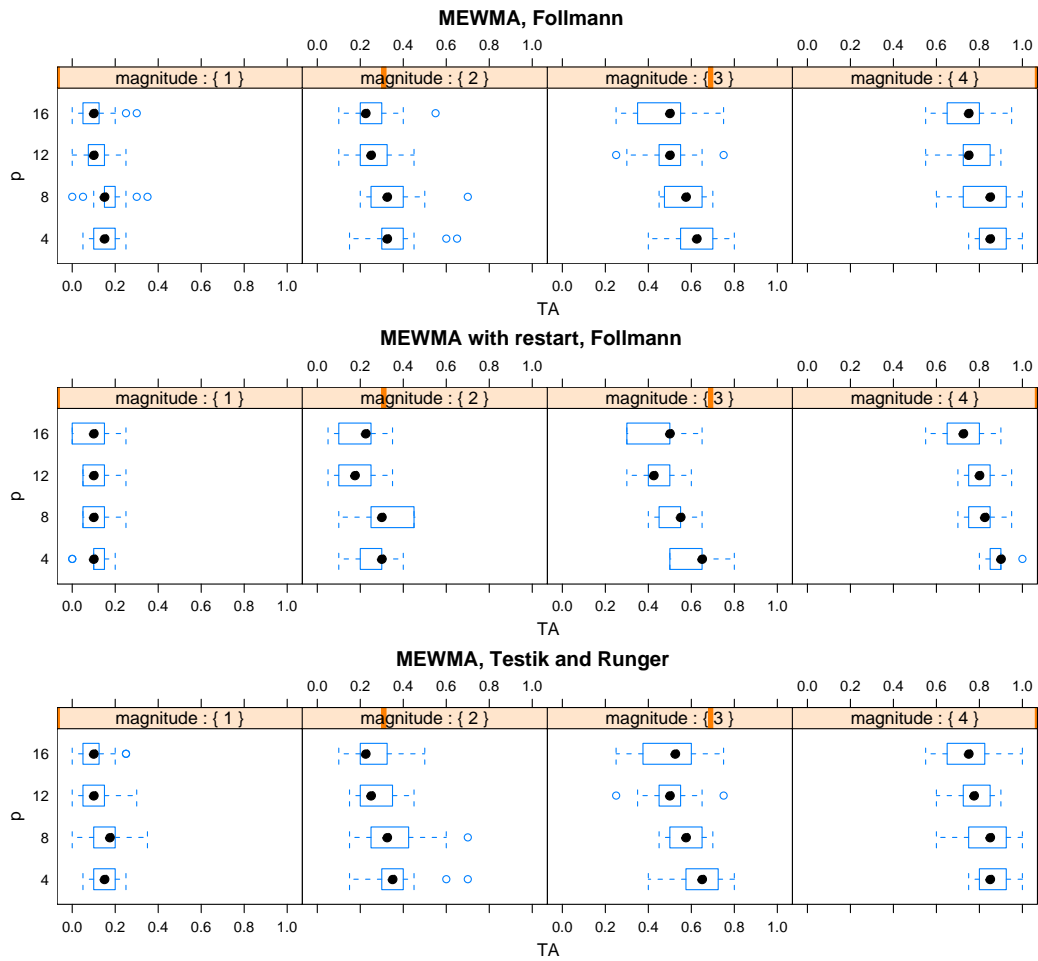


Figure 10: Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of spike magnitude

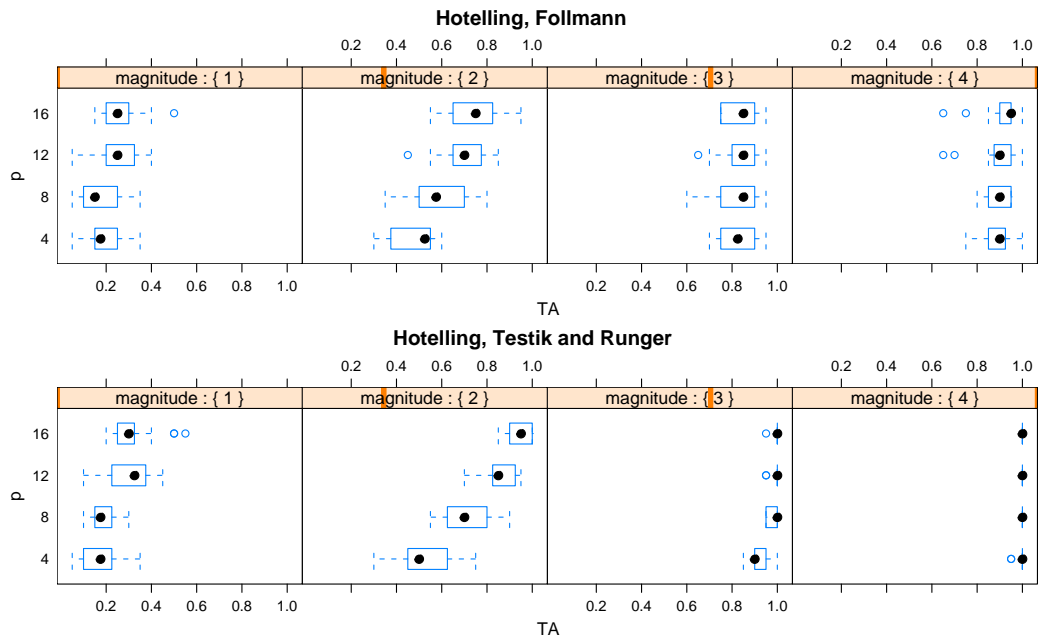


Figure 11: Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of spike magnitude when spike is injected into 25% of the series

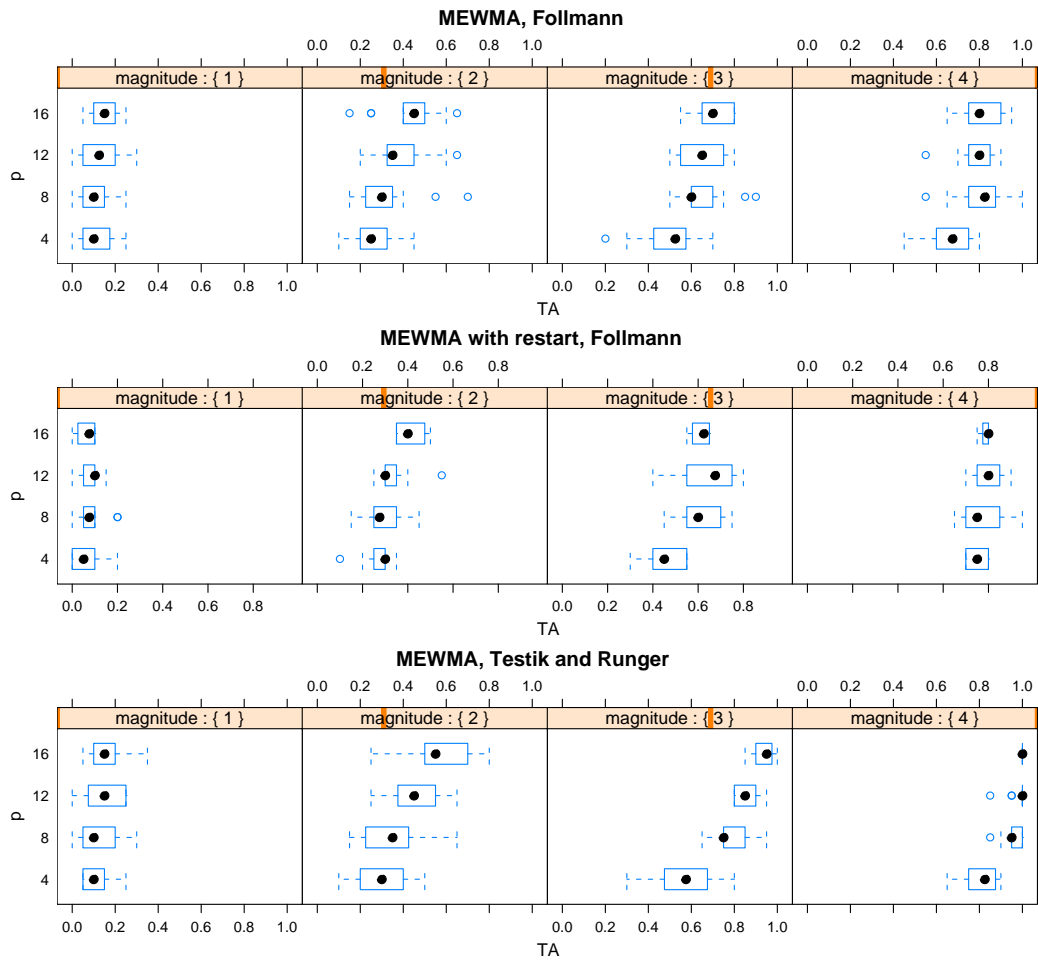


Figure 12: Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of spike magnitude when spike is injected into 25% of the series

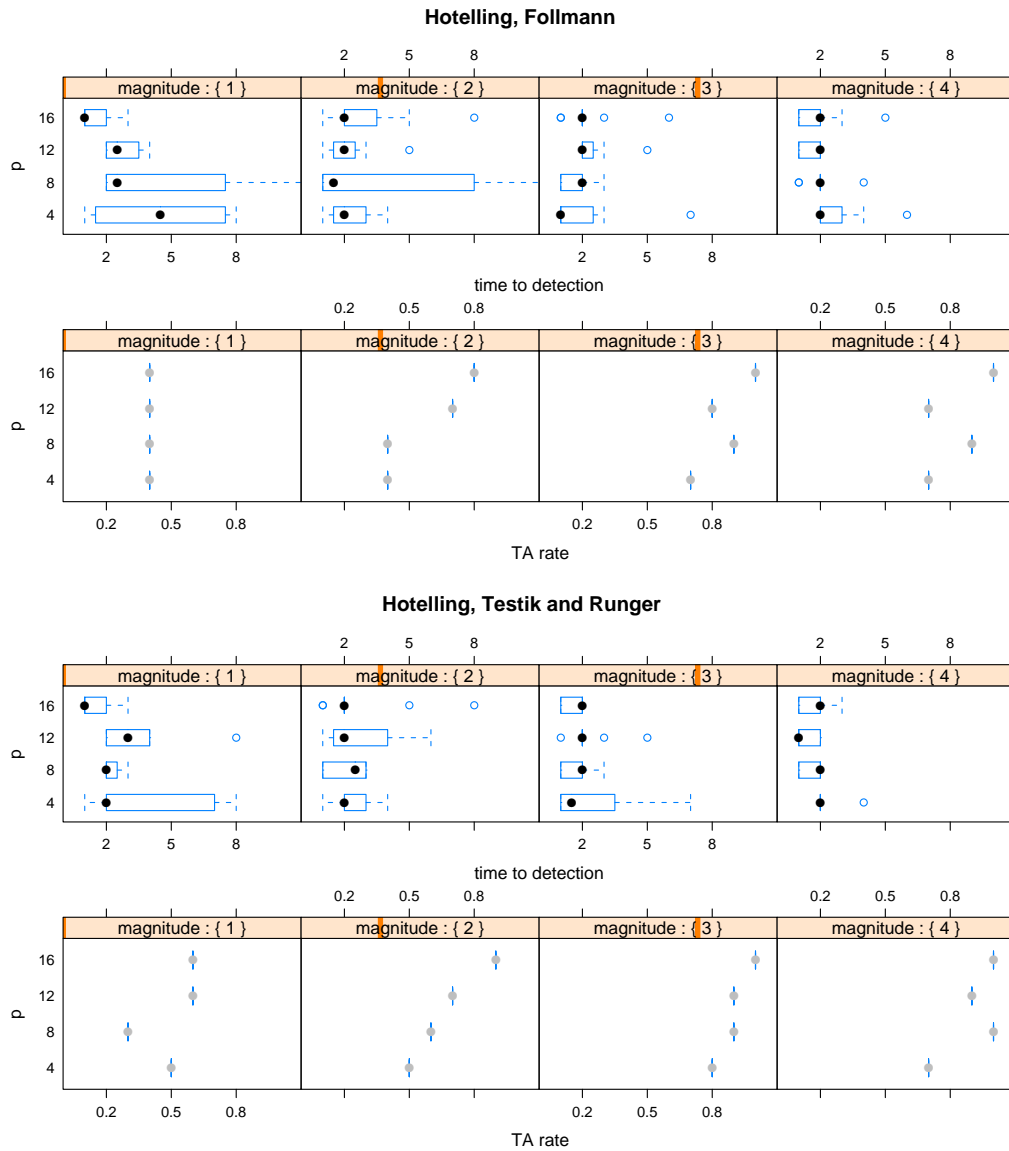


Figure 13: Distribution of true alert (TA) rate in directionally-sensitive Hotelling charts as a function of outbreak magnitude when the outbreak is injected into 25% of the series



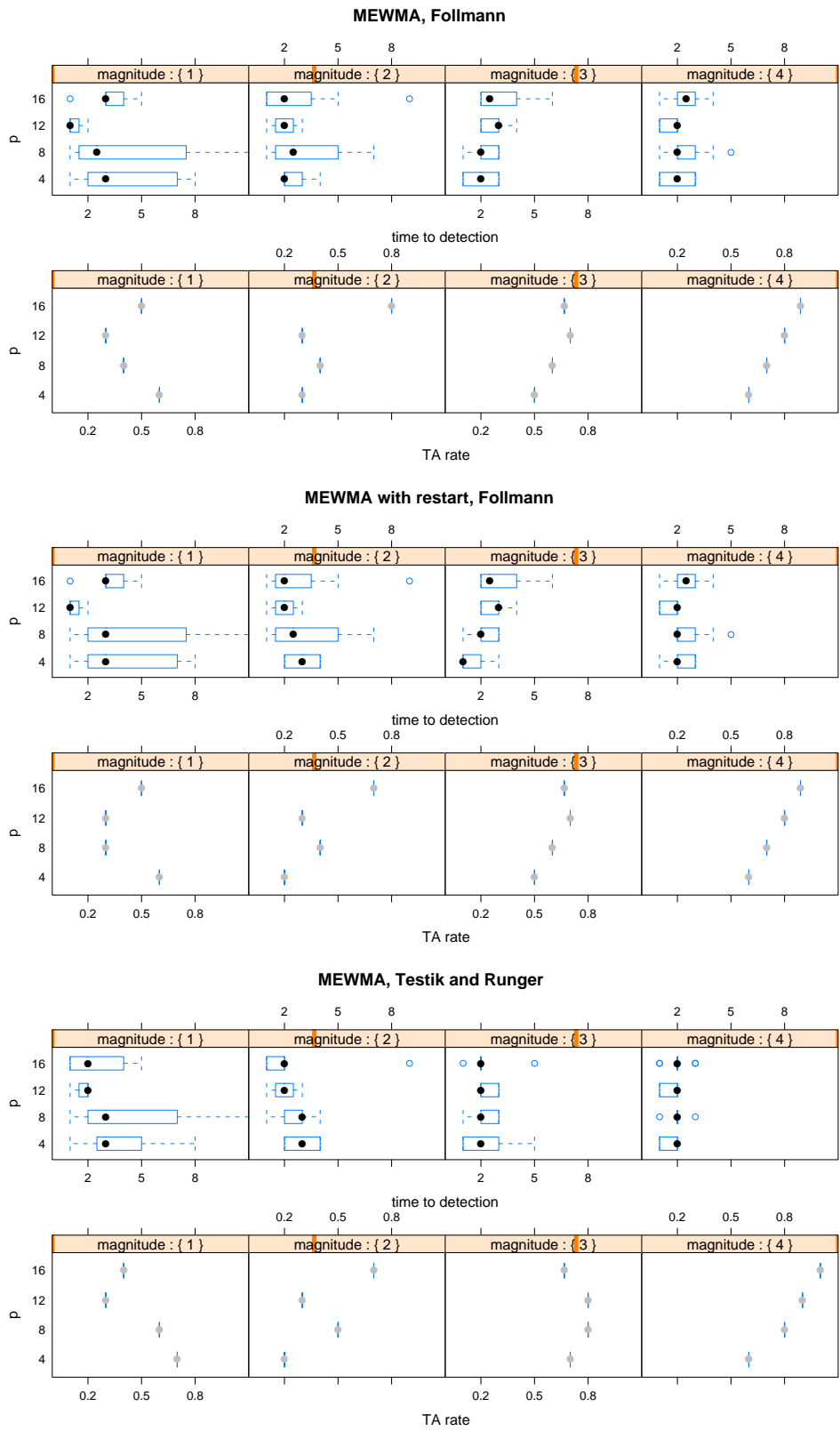


Figure 14: Distribution of true alert (TA) rate in directionally-sensitive MEWMA charts as a function of outbreak magnitude when the outbreak is injected into 25% of the series

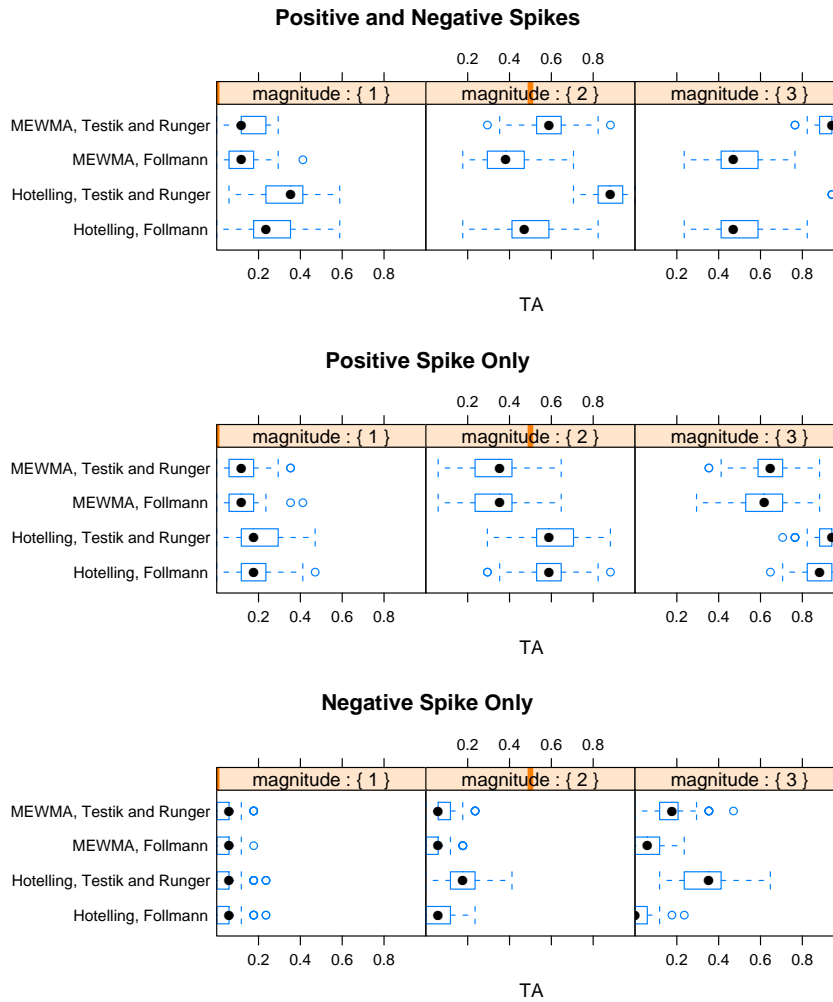


Figure 15: Distribution of true alert (TA) rate in directionally-sensitive charts, as a function of spike magnitude in the presence of increasing and decreasing spikes (top), increasing spikes only (middle) and decreasing spikes only (bottom)

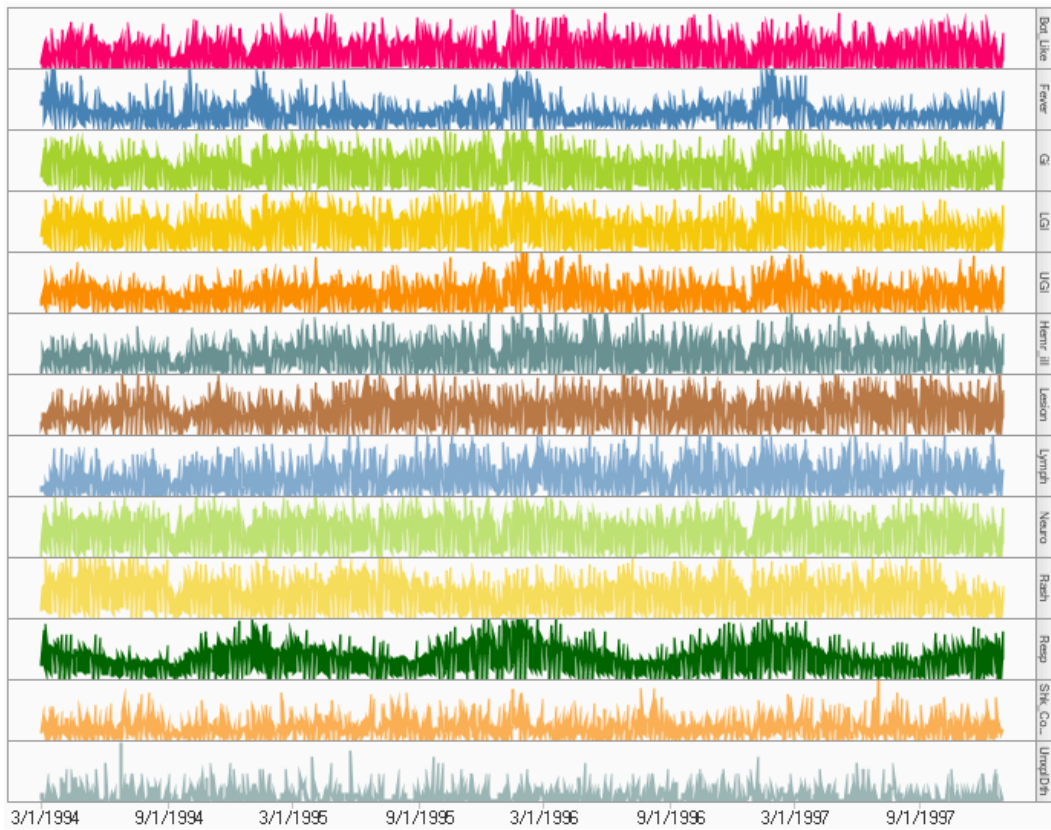


Figure 16: Authentic Data: Daily counts of chief complaints by patients arriving at emergency departments in a US city

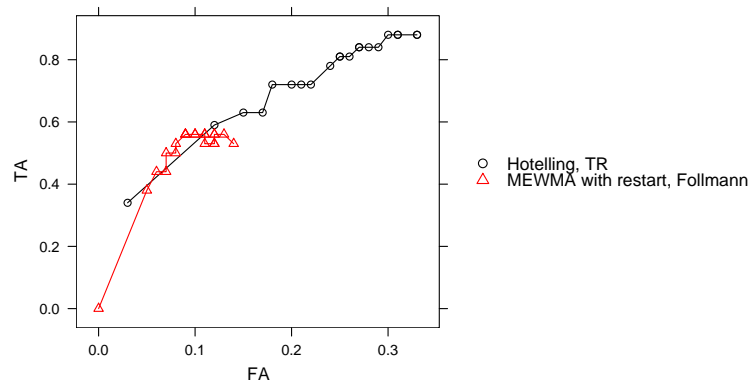


Figure 17: True vs. false alert rates for TR's Hotelling chart vs. Follmann's MEWMA chart with restarts

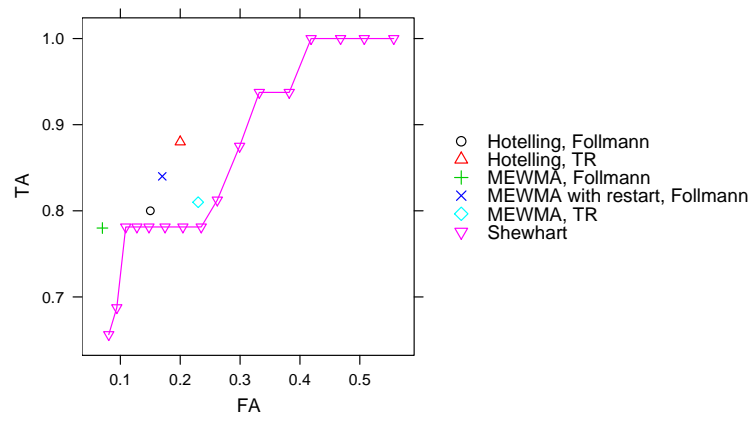


Figure 18: True vs. false alert rates; comparing multivariate control charts with multiple-univariate Shewhart charts

## List of Tables

1	The relationship between TA and FA rates . . . . .	39
2	Performance of the control charts on authentic data . . . . .	40

Table 1: The relationship between TA and FA rates

<i>Coef</i>	<i>Follmann's Hotelling</i>	<i>TR's Hotelling</i>	<i>Follmann's MEWMA</i>	<i>Follmann's MEWMA with Restart</i>	<i>TR's MEWMA</i>
$\widehat{\beta}_0$	<b>0.22</b> (0)	<b>0.33</b> (0)	<b>-0.31</b> (0)	-0.05 (0.27)	<b>-0.21</b> (0)
$\widehat{\beta}_1$	-0.62 (0.26)	-0.44 (0.39)	<b>5.35</b> (0)	<b>-3.78</b> (0)	<b>4.24</b> (0)
$\widehat{\beta}_2$	0.01 (0.64)	<b>-0.14</b> (0)	0.01 (0.21)	-0.02 (0.19)	<b>-0.11</b> (0)
$\widehat{\beta}_3$	<b>0.2</b> (0)	<b>0.2</b> (0)	<b>0.2</b> (0)	<b>0.25</b> (0)	<b>0.22</b> (0)
$\widehat{\beta}_4$	<b>0.01</b> (0)	<b>0.01</b> (0)	<b>0</b> (0)	<b>0.01</b> (0)	<b>0.01</b> (0)
Adj- $R^2$	0.58	0.58	0.71	0.7	0.75

Table 2: Performance of the control charts on authentic data

<i>Method</i>	<i>FA rate in the absence of outbreaks</i>	<i>FA rate in the presence of outbreaks</i>	<i>TA rate</i>
Hotelling, Follmann	0.16	0.15	0.80
Hotelling, Testik	0.16	0.20	0.88
MEWMA, Follmann	0.16	0.17	0.84
MEWMA with restart, Follmann	0.08	0.07	0.78
MEWMA, Testik	0.21	0.23	0.81