International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: http://www.tandfonline.com/loi/tprs20

A spatial variable selection method for monitoring product surface

Kaibo Wang, Wei Jiang & Bo Li

To cite this article: Kaibo Wang, Wei Jiang & Bo Li (2015): A spatial variable selection method for monitoring product surface, International Journal of Production Research, DOI: 10.1080/00207543.2015.1109723

To link to this article: http://dx.doi.org/10.1080/00207543.2015.1109723

Published online: 17 Nov 2015.

International Journal of Production Research, 2015
http://dx.doi.org/10.1080/00207543.2015.1109723

A spatial variable selection method for monitoring product surface

Kaibo Wang[a]*, Wei Jiang[b] and Bo Li[c]

[a]*Department of Industrial Engineering, Tsinghua University, Beijing, China;* [b]*Antai College of Economics and Management, Shanghai Jiaotong University, Shanghai, China;* [c]*School of Economics and Management, Tsinghua University, Beijing, China*

(*Received 24 March 2015; accepted 10 October 2015*)

Two-dimensional (2-D) data maps are generated in certain advanced manufacturing processes. Such maps contain rich information about process variation and product quality status. As a proven effective quality control technique, statistical process control (SPC) has been widely used in different processes for shift detection and assignable cause identification. However, charting algorithms for 2-D data maps are still vacant. This paper proposes a variable selection-based SPC method for monitoring 2-D wafer surface. The fused LASSO algorithm is firstly employed to identify potentially shifted sites on the surface; a charting statistic is then developed to detect statistically significant shifts. As the variable selection algorithm can nicely preserve shift patterns in spatial clusters, the newly proposed chart is proved to be both effective in detecting shifts and capable of providing diagnostic information for process improvement. Extensive Monte Carlo simulations and a real example have been used to demonstrate the effectiveness and usage of the proposed method.

Keywords: LASSO; semiconductor manufacturing; spatial clustering; statistical process control

1. Introduction

As a proven effective tool for quality control, statistical process control (SPC) has been widely utilised in diverse industries for special cause identification, removal and variation reduction (Montgomery 2009). Different control charts, ranging from univariate charts to multivariate charts, have been extensively studied in the literature. Recently, control charts for profiles are also drawing more and more attentions (Colosimo, Semeraro, and Pacella 2008). With the development of sensing technology, in an advanced process, process or product quality is sometime characterised by 2-D surfaces.

Taking the wafer fabrication process in semiconductor manufacturing as an example, wafer quality and yield are reflected or heavily influenced by spatial characteristics embedded on a 2-D surface. Figure 1 shows one wafer sample, with the site total indicator reading (STIR) reading of each site marked. The wafer is sliced from a silicon ingot, then processed and shipped to downstream manufacturers for integrated circuit (IC) fabrication. The manufacturing process will be introduced further in Section 7. Certain quality indices that reflect wafer thickness and flatness are important to process yields, therefore, should be closely measured and monitored. To evaluate wafer quality, the wafer is marked into different sites; the number of sites is determined by wafer size. Lin, Wang, and Yeh (2013) explained the detailed technical definition of different quality indices of wafers. Among others, STIR is one important quality index that indicates the flatness of each site on a wafer. The STIR of a wafer is affected its manufacturing processes. Changes in process condition may lead to sustained shifts on the wafer. Therefore, to better control wafer quality and prevent the propagation of defects to downstream production, it is important to monitor the 2-D STIR map and detect erroneous changes of critical quality characteristics during the production process.

The STIR readings form a 2-D data map structure with unique features. For examples, the readings are spatially distributed with certain neighbourhood structures; this phenomenon assigns additional positional information to each measure. As adjacent sites usually undergo similar treatment conditions, a process fault or change usually causes shifts in the format of clusters. Clustered changes in wafers with discrete readings have drawn the attention of many researchers (see, e.g. Hansen et al. 1997; Jeong, Kim, and Jeong 2008). In this work, we intend to detect process shifts by monitoring the continuous 2-D STIR map. A review of existing research on similar problems will be presented in Section 2.

However, a direct use of traditional SPC control charts is either difficult or inefficient due to the unique spatial characteristics the 2-D map data possess. The aim of this paper is to develop advanced SPC methods for monitoring 2-D surface. A spatial variable selection method is proposed jointly with the SPC monitoring algorithm for early detection of process failures so that defects can be prevented from spreading throughout the production process. In addition, a variable selection algorithm

*Corresponding author. Email: kbwang@tsinghua.edu.cn

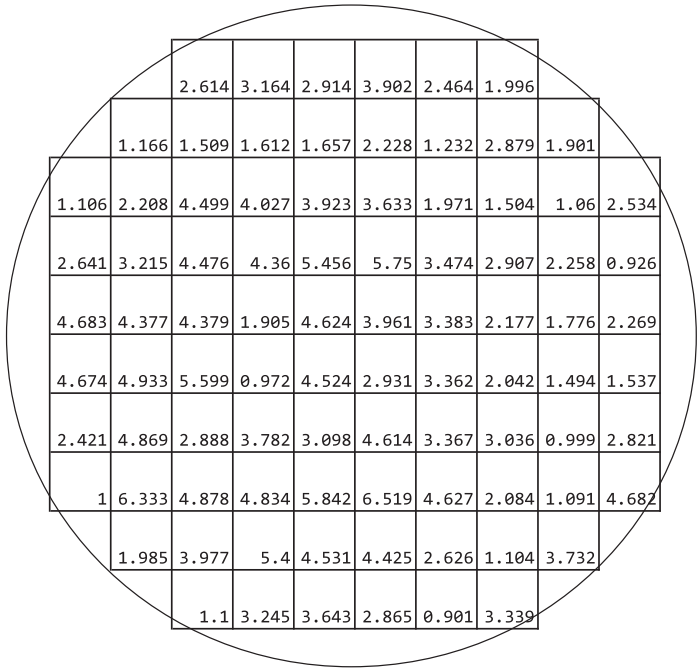


Figure 1. A sample STIR map.

is adapted to the above 2-D data to provide diagnostic information regarding the location, shape and size information if any alarms are triggered, thus facilitating root cause identification and removal.

The remaining of this paper is organised as follows. A literature review of current research on SPC methods for wafer quality monitoring is presented in the following section. In Section 3, we introduce a new control chart for monitoring process and product data in 2-D maps; a penalty-based method is employed to identify shift sites on the 2-D maps. Section 4 investigates the statistical performance of the proposed scheme for detecting spatially clustered shifts. Section 5 presents the design guidelines for practitioners on how to use the proposed method. Section 6 compares the proposed method with other alternatives including the industry benchmark. A real example is presented in Section 7 to illustrate the effectiveness of the proposed procedure. Finally, Section 8 concludes the paper with topics valuable for future research.

2. Review of SPC methods for 2-D surface data monitoring

In quality inspection, two types of data maps could be generated in the form of a 2-D map. For example, in die functionality testing in semiconductor manufacturing, a 2-D map with binary or integer data is usually observed. Hansen, Nair, and Friedman (1997) showed a wafer example of 300 chips on the top, with each chip being classified as either 'fail' or 'pass' through die function testing. Thus, a 2-D map of 0/1 is recorded for each wafer. Such 2-D maps share a similar structure as the one shown in Figure 1 except that the data are binary. In the literature, extensive studies on cluster identification using binary wafer maps has been reported (see, e.g. Hwang and Kuo 2007; Yuan and Kuo 2008; Liu et al. 2008). As shifts in the 2-D binary map are usually caused by process changes, SPC has also been employed to monitor such data maps. For example, Albin and Friedman (1989) studied the impact of clustered defects and proposed the use of the Neyman distribution to model and monitor the number of defects on a wafer. Hansen, Nair, and Friedman (1997) modelled a binary wafer map as an overlap of a normal map with defects randomly located and a map with defects shown in clusters, then developed testing statistics for checking the existence of spatial clusters by counting the number of neighbours of defective chips and the number of neighbours of nondefective chips. Hsu and Chien (2007) proposed a hybrid data mining approach to extraction shift patterns from binary 2-D data map. Jeong, Kim, and Jeong (2008) transformed a 2-D discrete data map using spatial correlogram for defect pattern identification. More recently, Chien, Hsu, and Chen (2013) developed an online algorithm for 2-D binary wafer map detection and classification.

Besides the 2-D discrete map, a 2-D continuous map, as the one shown in Figure 1, can also be collected. Different from the pass/fail discrete data, surface measurements such as STIR are continuous and carry richer information than discrete or binary defect data. Wang, Wu, and Wang (2015) analysed the control of a nanomanufacturing process, which produces

nanotube arrays in the form of a 2-D map. Such continuous measurements retain critical information of process steps, which can be utilised for root cause diagnosis if any shifts are detected.

There is a vast literature on modelling and analysis of spatial variation patterns. For example, Stine et al. (1997) proposed to decompose the variation of IC processes and devices into three levels: wafer-level, die-level and wafer-die interaction. Jeong, Lu, and Wang (2006) proposed a wavelet-based procedure for functional data analysis, which is a more general type of data with particular spatial structures. Reda and Nassif (2009) assumed that the spatial measurements collected from wafer testing follow a multivariate normal distribution. Cheng et al. (2011) studied the die-level variation, and suggested that spatial variation comes from deterministic across-wafer variation, while purely random spatial variation is not significant. Bao et al. (2014) modelled wafer variation using a hierarchical model. These works have provided insights about the variation a wafer has and are expected to benefit the practice with a better understanding of their processes. However, such analysis is carried out targeting at the usual variation a manufacturing process exhibits, which is what the term 'common-cause variation' represents in SPC. When a process goes out of control due to process shifts, component failures or other unexpected changes, 'special-cause variation' emerges.

SPC is a quality control technique designed to monitor process status, detect process shifts and provide diagnostic information if needed. Although SPC has been used widely in diverse industries and proven effective for variation reduction, SPC methods for wafer monitoring using 2-D continuous data map is still vacant in the literature. In the current semiconductor industry practice we have experienced, the maximum STIR value on each wafer is used for monitoring. This strategy obviously ignores information embedded in the 2-D map, which should be utilised to enhance process monitoring for detecting spatial clusters and diagnosing root causes.

Product surface data could be monitored using multivariate SPC charts if the measures at each site are considered a random variable and the 2-D map is represented by a multivariate random vector. However, traditional multivariate SPC methods don't consider spatial characteristics the data have and may lose efficiency when monitoring and identifying spatial clusters in 2-D maps. Moreover, the vector often has very high dimensions since the number of sites on a 2-D map increases quadratically with its physical dimension. Therefore, SPC methods without dimension reduction are rendered useless in practice (Jiang and Tsui 2008).

Recently, novel multivariate SPC methods with dimension reduction capability have been developed to alleviate the high dimensionality problem. Wang and Jiang (2009) and Zou and Qiu (2009) independently proposed to conduct variable selection in a multivariate vector and monitor only specific directions identified by the variable selection algorithms. Jiang, Wang, and Tsung (2012) extended the work of Wang and Jiang (2009) to an EWMA-based framework. The authors have shown that the variable selection-based multivariate control charts are appealing, especially when the dimension of observation vectors is very high. However, these variable selection-based control charts are generic; the unique while important spatial information in the 2-D STIR map is not taken into consideration. Such information may be helpful in identifying spatial areas that result in clustered defects. Therefore, in this work, we propose to extend the variable selection-based SPC methods to monitor 2-D surface data. Potential defect clusters are first screened out via spatial variable selection while neighbourhood relationships between adjacent sites are considered when locating spatial clusters. A control chart statistic is then designed to monitor the identified cluster(s). By taking advantages of the spatial relationships, it is expected that the proposed multivariate SPC method performs more efficiently in process monitoring and diagnosis for improving wafer quality.

3. A fused LASSO-based control chart

To characterise the 2-D surface map shown in Figure 1, we treat the value at each site as a random variable, and thus represent the whole map using a high-dimensional random vector $\boldsymbol{y}_t$, where t represents time order or wafer index. Without loss of generality, we assume that when the process is in control, the map has no cluster of mean shifts and each site is standardised to follow a normal distribution with mean zero and unit variance. In addition, all sites are assumed independent. That is, $\boldsymbol{y}_t \sim MN(\boldsymbol{0}, \boldsymbol{I})$ where $\boldsymbol{I}$ is an identity matrix. When the manufacturing process that produces the wafer deteriorates, mean shifts are usually seen in multiple adjacent sites. That is, $\boldsymbol{y}_t \sim MN(\boldsymbol{\mu}, \boldsymbol{I})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)'$ is a p-dimensional vector with nonzero elements indicating shift magnitude. An identity covariance matrix is assumed in this work while it is not difficult to generalise the proposed method to the case with a general covariance matrix.

In order to monitor the mean changes in $\boldsymbol{y}_t$, Jiang and Tsui (2008) developed a likelihood ratio test framework for the following hypotheses,

$$\begin{cases} H_0 : \boldsymbol{\mu} = \boldsymbol{0} \\ H_1 : \boldsymbol{\mu} \neq \boldsymbol{0} \end{cases}.$$

Since the mean vector is not specified in the alternative hypothesis, the generalised likelihood ratio test (GLRT) can be developed by contrasting the likelihood functions under the null and alternative hypotheses. The null hypothesis is rejected

when

$$\boldsymbol{y}_t^T \boldsymbol{y}_t - S_t > h \tag{1}$$

holds, where h is the threshold that controls the type I error probability α and

$$S_t = \min_{\boldsymbol{\mu} \neq \mathbf{0}} \left(\boldsymbol{y}_t - \boldsymbol{\mu}\right)^T \left(\boldsymbol{y}_t - \boldsymbol{\mu}\right) \tag{2}$$

provides the maximum likelihood estimate of the mean parameter $\boldsymbol{\mu}$ under the alternative hypothesis. It can be shown that the minimum is attained when $\hat{\boldsymbol{\mu}}_t = \boldsymbol{y}_t$ and the GLRT statistic reduces to the well-known Hotelling's T^2 statistic, which measures the distance between $\boldsymbol{y}_t$ and $\mathbf{0}$.

The above-mentioned derivation shows that the Hotelling's T^2 statistic can be viewed as a two-step procedure: first, estimating the mean shift under the alternative hypothesis using $\hat{\boldsymbol{\mu}}_t = \boldsymbol{y}_t$; second, testing the hypotheses along the direction of $\hat{\boldsymbol{\mu}}_t$. As Wang and Jiang (2009) noted, it is inappropriate to take $\boldsymbol{y}_t$ as an estimate of mean shift since all the components of $\boldsymbol{y}_t$ are generally nonzero, which implicitly imply all components shift simultaneously. Different methods can be used to estimate the mean shift under the alternative hypothesis. For example, in the adaptive T^2 chart developed by Wang and Tsung (2008), the authors used a model-free EWMA statistic to estimate the mean shift using historical observations. Once an estimate of $\boldsymbol{\mu}$ is obtained, the process is monitored using a directionally variant chart. Alternatively, Wang and Jiang (2009) assumed that assignable causes usually affect a few rather than all variables at the same time in a multivariate process. They suggested using variable selection techniques to first find suspected variables and then integrate such information with the monitoring statistic for multivariate process control.

The variable selection algorithm used in Wang and Jiang (2009) and Jiang, Wang, and Tsung (2012) is generic, which does not utilise any domain knowledge of the manufacturing process. It has been shown to be effective to facilitate the SPC monitoring method in Equation (1) with known shift information. That is, if an estimate of the mean shift can be known, the charting algorithm can be tuned to be more effective with respect to the suspected shifts. In monitoring a 2-D data map, since assignable causes may result in defects in spatial clusters of adjacent sites, it is expected that the spatial characteristics may provide particular assistance for capturing mean shifts in clusters, thus help with improved charting performance. Therefore, we here propose to estimate clustered mean shifts based on a penalised version of Equation (2).

3.1 *Statistical methods for monitoring 2-D surface*

In order to efficiently identify out-of-control clusters in a 2-D surface, we consider a modification of the maximum likelihood estimate in the optimisation problem (2) to locate adjacent sites with potential shifts. Therefore, for efficient control chart design, an estimate of process shifts is expected to have the following properties.

First, it should force the mean of in-control sites to be zero. In practice, we believe that assignable causes only result in mean shifts in certain sites and most sites remain in control under the assignable cause of variations. For example, in the wafer lapping process illustrated in Lin and Wang (2011), impurities in the lapping slurry may cause scratches on wafer surface. That is, quality shifts on a 2-D wafer surface turned to be sparse. The optimisation solution is expected to set the mean of in-control sites to zero automatically so that we only need to focus on the sites that have nonzero mean estimates. This will help reduce the dimension that needs to be monitored.

Second, the mean estimate can identify out-of-control sites in clusters. It is already recognised that in wafer applications, defects and shifts tend to occur in clusters. As adjacent areas on a wafer are processed under similar conditions, once an assignment root cause occurs, it usually damages sites that are close to each other. Therefore, the algorithm should take such engineering knowledge into consideration when searching for suspected sites, although the size and location of clusters are all unknown. To achieve the above objectives, we start from the objective function in Equation (2) and modify its solution path by adding new penalties.

When monitoring the mean of a multivariate data stream, Wang and Jiang (2009) and Zou and Qiu (2009) proposed to use the following modified objective function:

$$S_t^{(1)} = \min_{\boldsymbol{\mu} \in \Omega_1} \left[(\boldsymbol{y}_t - \boldsymbol{\mu})^T (\boldsymbol{y}_t - \boldsymbol{\mu}) + \sum_{j=1}^{p} g_{\lambda_j}(|\mu_j|) \right], \tag{3}$$

where $g_{\lambda_j}(\cdot)$ is a penalty for achieving certain patterns in μ and λ_j's are tuning parameters controlling the strength of penalisation. If $g_{\lambda_j}(|\mu_j|) = \lambda_j |\mu_j|$, Equation (3) is a function with LASSO-type L_1-penalty, which was proposed by Tibshirani (1996). If $g_{\lambda_j}(|\mu_j|) = \lambda_j I(|\mu_j| \neq 0)$, the last term in Equation (3) is an L_0-penalty that constrains the weighted number of nonzero coefficients in $\boldsymbol{\mu}$. Both L_0- and L_1-penalties can help to obtain sparse estimators and hence make the

resulting monitoring more concentrative. Wang and Jiang (2009) used the L_0-penalty and Zou and Qiu (2009) used an adaptive L_1-penalty to select suspected variables when monitoring multivariate processes.

However, the sparse property of the usual L_0- and L_1-penalties does not utilise any positional information. When monitoring 2-D surface data, as aforementioned, shift sites with nonzero mean values tend to form clusters. In order to attain sparse but clustered estimator, Tibshirani et al. (2005) proposed a fused LASSO algorithm to further penalise the L_1-norm of both coefficients and their successive differences. Assuming $\boldsymbol{y}_t$ is a sample taken along a 1-D line, the corresponding objective function is expressed as follows:

$$S_t^{(2)} = \min_{\boldsymbol{\mu}\in\Omega_1} \left[(\boldsymbol{y}_t - \boldsymbol{\mu})^T (\boldsymbol{y}_t - \boldsymbol{\mu}) + \lambda_1 \sum_{j=1}^{p} |\mu_j| + \lambda_2 \sum_{j=1}^{p-1} |\mu_j - \mu_{j+1}| \right].$$

The newly added difference penalty (λ_2) helps prune adjacent mean estimates towards each other, thus forming segments of zeros or nonzero but equal values.

Tibshirani et al. (2005) only considered the 1-D case in which $\boldsymbol{y}_t$ is a vector; it cannot be used directly to the 2-D data shown in Figure 1. Friedman et al. (2007) extended the fused LASSO to the 2-D case and connected it with the widely used 'total variation denoising' procedure in signal processing (Rudin et al. 1992), particularly image reconstruction. Hoefling (2010) studied the 2-D fused LASSO more extensively and developed a fast path algorithm for implementation, which is briefly discussed in the Appendix. A general spatial fused LASSO method used the following penalised loss function (Hoefling 2010),

$$S_t^{(3)} = \min_{\mu\in\Omega_1} \left[(\mathbf{y}_t - \mu)^T (\mathbf{y}_t - \mu) + \lambda_1 \sum_{j=1}^{p} |\mu_j| + \lambda_2 \sum_{(s,t)\in E, s<t} |\mu_s - \mu_t| \right], \tag{4}$$

where E is a set of neighbourhood clusters consisting of adjacent sites in the map. That is, the modified penalty can be used to force neighbouring sites to have similar magnitude of mean shifts. The definition of the set E is quite flexible. In the examples presented in Friedman et al. (2007) and Hoefling (2010), they considered the immediate left, right, upper and lower sites as neighbours of a focused site. In our example, we also accommodate the immediate upper left, upper right, lower left and lower right sites as neighbours. That is, each site has eight neighbours (unless the site is close to boundary). We denote the solution of Equation (4) as $\tilde{\boldsymbol{\mu}}_t$ in the following discussion.

The solution of Equation (4) has the following properties. First, by properly defining neighbours, the added fusion penalty (λ_2) can help forcing equal shift magnitude among neighbours, which builds the foundation for separating potential out-of-control and in-control clusters and forming clusters of different shift magnitude. Second, the LASSO penalty (λ_1) helps shrinking small but nonzero elements to zero, thus separates the in-control sites from the suspects, leaving potential out-of-control clusters stand out.

It is important to note that the above procedure is fully data-driven in finding mean shift clusters. Without the need to specify cluster size and location, it can locate suspected clusters automatically. In general, the two parameters λ_1 and λ_2 control the size and significance of the clusters identified. Intuitively, the larger the values of λ_1, the fewer number of clusters is identified; the larger the values of λ_2, the smaller size the identified clusters are. Since our objective is not only to identify shift clusters, but also to monitor high-dimensional processes with fewer variables, appropriate choices of these parameters for SPC monitoring will be discussed in the next section.

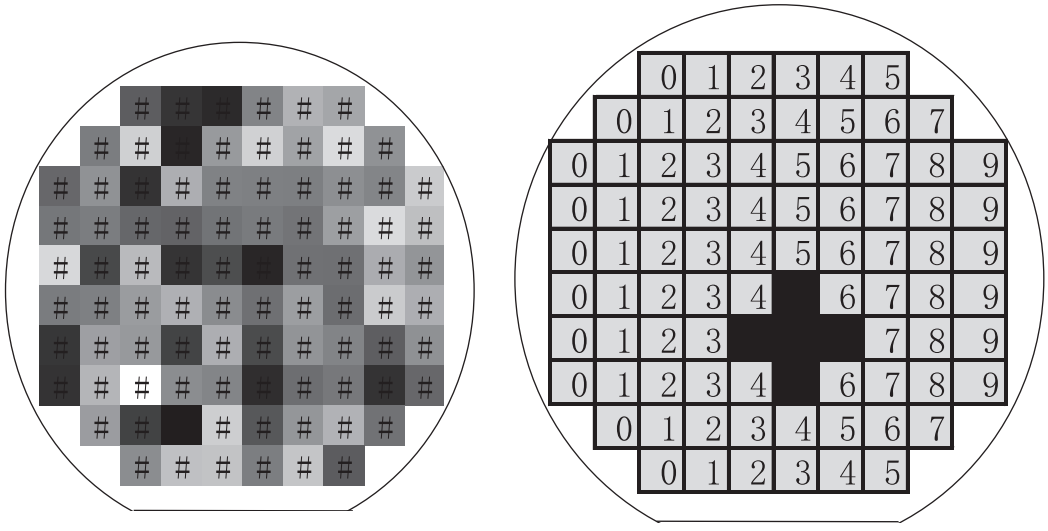
3.2 *A fused LASSO-based control chart*

Based on the fused LASSO algorithm, we can successfully obtain sparse and clustered estimates of site means. We now develop a multivariate control chart to detect the significance of such estimates. We assume that the solution of Equation (4) is the suspected shifts of the 2-D surface. It is known that in multivariate process monitoring, directionally variant chart is more efficient than the conventional T^2 chart given the shift direction information (Wang and Tsung 2008; Wang and Jiang 2009). Therefore, we suggest monitoring the following statistic to detect any clustered mean shifts,

$$\boldsymbol{\Lambda}_t = \tilde{\boldsymbol{\mu}}_t^T \boldsymbol{y}_t > h \tag{5}$$

It can be shown that this statistic is equivalent to that in Equation (1). If out-of-control clusters are estimated accurately, the above control chart is expected to perform better than the Hotelling's T^2 chart. As this chart is derived based on the fused LASSO variable selection algorithm, we call this chart the fused LASSO-based multivariate statistical process control (MSPC) chart (denoted as FL-MSPC hereafter).

By integrating with the fused LASSO variable selection method, the FL-MSPC chart possesses the following features. First, the variable selection algorithm is helpful in reducing dimensionality. Given an observed sample with a large number



(a) In-control surface map (b) Shift pattern

Figure 2. A simulated surface sample and shifted cluster.

of sites, the potential out-of-control sites are identified automatically by the fused LASSO algorithm. Second, the spatial neighbourhood information among sites is considered by the fused LASSO method and retained in the solution $\tilde{\boldsymbol{\mu}}_t$. Third, the result of variable selection, $\tilde{\boldsymbol{\mu}}_t$, is time varying and helpful in fault diagnosis. If any out-of-control signal is triggered, we can locate the responsible area by checking nonzero components in $\tilde{\boldsymbol{\mu}}_t$. Those sites with nonzero estimates should be responsible for the signal. Therefore, the FL-MSPC chart is capable of achieving process monitoring and diagnosis in a unified approach.

In the following, we shall first study the performance of the FL-MSPC chart for spatial cluster identification and monitoring. We then compare the performance of the FL-MSPC chart with other candidates that can be used for 2-D surface. Finally, an industrial example from a wafer preparation process is analysed for spatial fault detection.

4. Performance analysis of the FL-MSPC chart

One of the most important metrics to assess a control chart is average run length (ARL), which is the average number of observations collected (or run steps) until the first out-of-control signal triggered by the control chart. In this section, we first study the sensitivity of the fused LASSO algorithm in identifying spatial clusters. Then choices of parameters for designing the FL-MSPC chart are discussed.

4.1 *Performance of the fusion and LASSO penalties*

In this section, we still follow the wafer example. Generally, the dimension of a wafer map is determined by wafer size. For illustration, the wafer map shown in Figure 1 is used to configure the FL-MSPC chart. In this example, the wafer surface is overlaid by 10 by 10 grid lines. By removing blocks around 4 corners, we have 88 sites on each wafer that have STIR measurements for surface monitoring. In the fused LASSO algorithm, each inner cell has eight neighbour sites around. For edge or corner sites, some of their neighbours are cut off. If the wafer map is different from Figure 1, similar analysis can be performed.

In the following, we will study the performance of the fused LASSO algorithm and confirm its properties in dimension reduction for cluster identification. For performance analysis, we assume that when the wafer is in control, each site is independent of other sites and follows a standardised normal distribution. When a process shift occurs, it affects the mean of a cluster of adjacent sites. For illustration, we randomly generate a sample of in-control surface with each site independently distributed, as shown in Figure 2(a). The grayscale indicates each site's STIR value. It is seen that since the process is in control, there is no apparent spatial patterns on the surface. The shift pattern in Figure 2(b) is then added to the surface sample. The fused LASSO algorithm is applied to verify its performance in shift pattern identification.

4.1.1 *When shift magnitude increases*

Figure 3(a)–(c) shows the performance of the fused LASSO algorithm with λ_1=0.1 and $\lambda_2 = 0.2$ for identifying the cluster in Figure 2 when the shift size δ changes from 1 to 5. The corresponding sites and clusters identified by the fused LASSO algorithm is shown in shaded areas (the darker, the higher STIR values the corresponding sites have). The grayscale of each site reflects the variable selection solution obtained via Equation (4). Those sites with the same colour are grouped by the fusion term in the model. When the shift size increases, we can see that the shift cluster becomes more apparent. Thus, it is exemplified that the fused LASSO algorithm is capable of identifying spatial clusters with sites of approximately equal shift sizes. Meanwhile, nonshift sites can also be identified, grouped and set to zero by the LASSO penalty.

4.1.2 *The sensitivity of parameters for spatial cluster identification*

Next, we vary the value of the two parameters λ_1 and λ_2 to see their effects for cluster identification. From the model in Equation (3), the LASSO coefficient λ_1 is used to force certain nonzero components to zero. A large value tends to generate more zeros in $\hat{\boldsymbol{\mu}}$. In the extreme case, a sufficiently large value of λ_1 will turn all elements in $\hat{\boldsymbol{\mu}}$ to zero, while a zero value of λ_1 makes the solution lose the sparse property and very likely no zeros can be found in $\hat{\boldsymbol{\mu}}$. On the other hand, the fusion coefficient λ_2 determines how stairs are merged and clusters are formed. A large value of λ_2 tends to put more emphasis on cluster edges and force less clusters, while a small value of λ_2 tends to ignore the boundary effect of clusters. As a special case, when $\lambda_2 = 0.0$, cluster boundaries are totally ignored and the fused LASSO algorithm reduces to the regular LASSO algorithm, which does not consider any spatial neighbourhood relationship. Consequently, the identified clusters may randomly scatter over the map.

For the cluster shift pattern assigned in Figure 2, Figure 3 shows cases when changing the penalty parameters or shift sizes. It is easy to see that when λ_1 increases, the boundary between chosen and unchosen sites gradually stands out; the cluster becomes clearer when other sites are set to zero. When λ_1 becomes extremely large, it is possible that all sites are set to zero. It is also evident that the added fusion penalty is helpful in removing noise and forming clusters. Therefore, the shift pattern in (g)–(l) is mostly more prominent than that in (d)–(f).

The fused LASSO algorithm is briefly outlined in the Appendix. In the fusion step, clusters are formed by penalising boundaries among clusters. In the LASSO step, a soft-thresholding operation is applied to the fusion solution to shrink shift size and remove sites with small shifts. The above observations drawn from Figure 3 are consistent with our understanding of the algorithm.

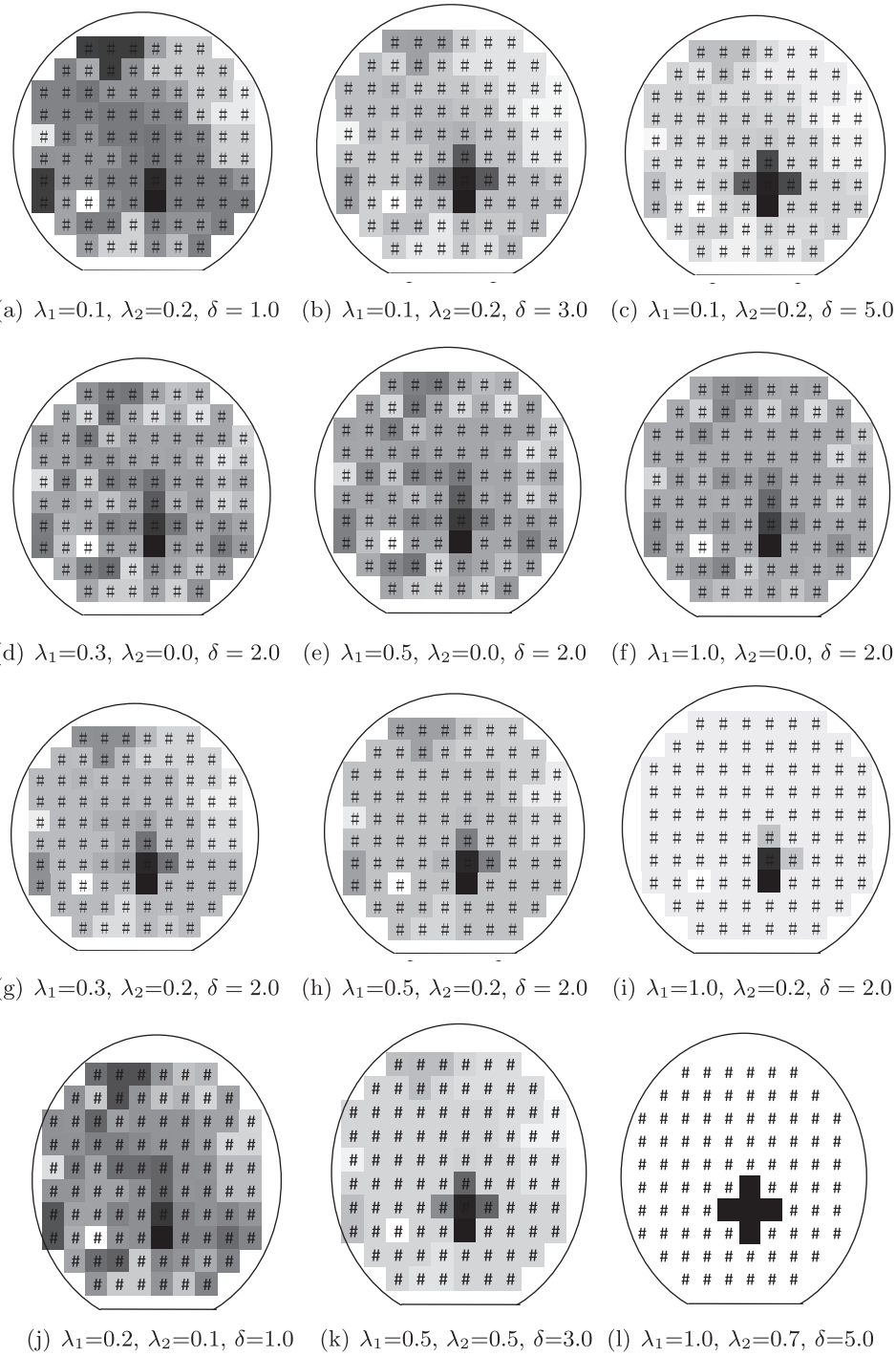
4.2 *Run length performance of FL-MSPC chart*

To illustrate the performance of the FL-MSPC chart for detecting spatial clusters on a 2-D map, we use the above wafer example and simulate different failure patterns, namely Patterns A, B, C, D, E and F in Figure 4 for study. Note that the cluster patterns in Figure 4 are designed to cover both single-site shifts (Pattern A) and large cluster shifts (Patterns C and D), multiple cluster shifts (Pattern E) and complex patter shifts (Pattern F). When any shifts occur, it is assumed that all changed sites suffer from the same magnitude of sustained shifts. The shift magnitude varies from 0.05 to 5.0 to cover small, moderate and large sizes, which is common considered in SPC literature.

It should be noted that the FL-MSPC is location invariant, that is, its performance is not affected by the location of shift clusters, as long as the shape of shift clusters and their neighbour relationship remain unchanged. This happens since the chart statistic does not consider sequence of the variables in the vector. Moving cluster locations is equivalent to moving a block of variables together in the vector. So, the charting performance will not be affected.

To illustrate the effect of the penalty parameters λ_1 and λ_2, we define relative efficiency (RE) of a FL-MSPC chart as the ratio of ARL's between the conventional T^2 chart and the corresponding FL-MSPC chart for detecting the same cluster of mean shifts. A value of RE larger than one implies superior performance of the FL-MSPC chart compared to the conventional T^2 chart.

We first discuss the case when using only one penalty parameter for variable selection in Equation (4). Figure 5 presents the RE values of different FL-MSPC charts with only one penalty parameter. Note that when $\lambda_2 = 0.0$, the fused LASSO algorithm reduces to the regular LASSO algorithm and the FL-MSPC chart is analogous to the variable selection multivariate statistical process control (VS-MSPC) chart proposed by Wang and Jiang (2009). As an example, Figure 5(a) plots of the RE values of the FL-MSPC chart for detecting the six types of clusters when $\lambda_1 = 0.2$ and $\lambda_2 = 0.0$. It is easy to see that for a single-site mean shift, the FL-MSPC chart outperforms the T^2 chart for detecting shifts of almost all clusters with moderate shifts. This agrees with our expectation that the LASSO algorithm can improve the shift detection through dimension reduction. The performance of the FL-MSPC is especially prominent for large and simple clusters, such as Patterns



(a) λ_1=0.1, λ_2=0.2, $\delta = 1.0$ (b) λ_1=0.1, λ_2=0.2, $\delta = 3.0$ (c) λ_1=0.1, λ_2=0.2, $\delta = 5.0$

(d) λ_1=0.3, λ_2=0.0, $\delta = 2.0$ (e) λ_1=0.5, λ_2=0.0, $\delta = 2.0$ (f) λ_1=1.0, λ_2=0.0, $\delta = 2.0$

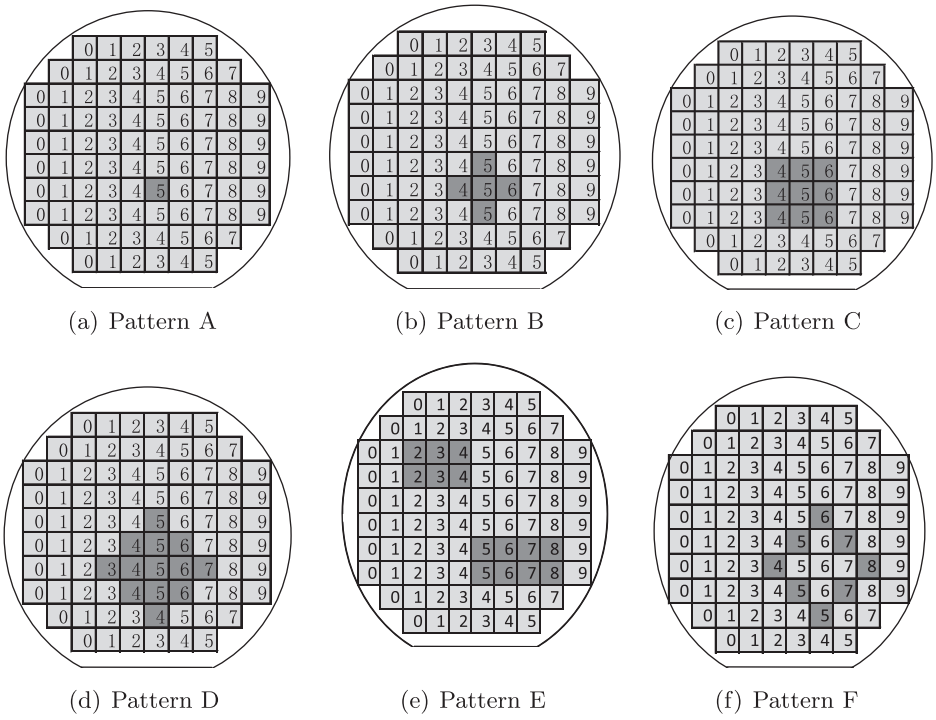
(g) λ_1=0.3, λ_2=0.2, $\delta = 2.0$ (h) λ_1=0.5, λ_2=0.2, $\delta = 2.0$ (i) λ_1=1.0, λ_2=0.2, $\delta = 2.0$

(j) λ_1=0.2, λ_2=0.1, δ=1.0 (k) λ_1=0.5, λ_2=0.5, δ=3.0 (l) λ_1=1.0, λ_2=0.7, δ=5.0

Figure 3. Clusters selected by the fused LASSO algorithm. (a)–(c): fixed $\lambda_1 = 0.1$ and $\lambda_2 = 0.2$, varying shift size δ; (d)–(f): fixed $\lambda_2 = 0.0$ and $\delta = 2.0$, varying λ_1; (g)–(i): fixed $\lambda_2 = 0.2$ and $\delta = 2.0$, varying λ_1; (j)–(l): varying λ_1, λ_2 and δ.

B, C, and D. For complex shift patterns E and F, the FL-MSPC chart also improves the charting performance, although not as significant as other patterns.

On the other hand, Figure 5(b) plots the RE values of the FL-MSPC chart for detecting the six types of clusters when $\lambda_1 = 0.0$ and $\lambda_2 = 0.2$. It is easy to see that the FL-MSPC chart significantly outperforms the conventional T^2 chart if the cluster is not too small. The larger size the cluster has, the more significant the RE improvement is for the FL-MSPC chart, especially when the shift magnitude is small. This indicates that the fusion penalty is very effective in identifying clusters



(a) Pattern A (b) Pattern B (c) Pattern C

(d) Pattern D (e) Pattern E (f) Pattern F

Figure 4. Simulated cluster patterns.

and consequently improving the change detection performance. When the shift magnitude is large, both T^2 and FL-MSPC methods can detect the change very quickly, so the advantage of fusion penalty diminishes. For detecting mean shifts in a single site, the fusion penalty fails to identify mean shifts with small magnitude due to random errors but still effectively detect shifts of large magnitude, compared to the conventional T^2 method.

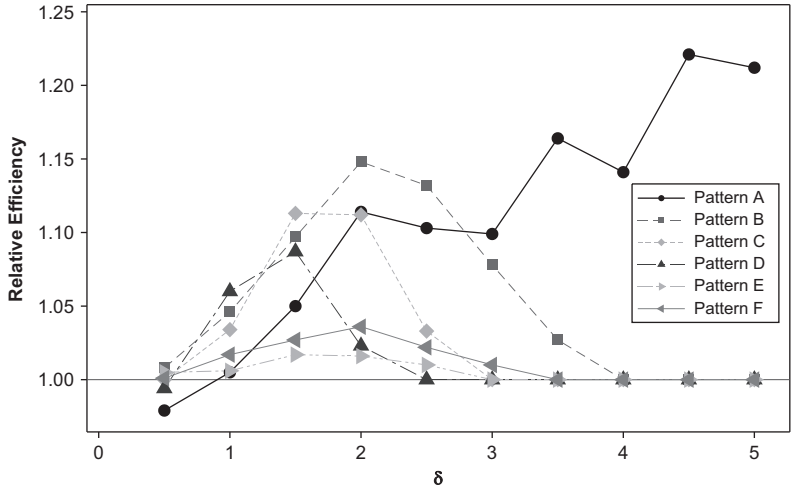
It is important to note that the RE values of the FL-MSPC chart over the T^2 chart are more prominent in Figure 5(b) than those in (a), especially for large clusters of mean changes. For example, the FL-MSPC chart with only the fusion penalty $\lambda_2 = 0.2$ improves over the conventional T^2 chart by more than 50% for detecting a mean shift of cluster pattern C and magnitude 1.5, whereas the FL-MSPC chart with only L_1-penalty $\lambda_1 = 0.2$ improves only 2%. Even for detecting a single site of clusters, the former FL-MSPC chart significantly outperforms the latter FL-MSPC chart in detecting shifts of large magnitudes, although the former performs worse in small magnitudes. This implies that the fusion penalty is in fact effective for dimension reduction in the case of large magnitudes, although it does not aim to reduce the dimension. Nevertheless, the fusion penalty suffers from inevitable randomness when the shift magnitude is too small and renders disadvantageous of the corresponding FL-MSPC method.

5. Design guidelines for FL-MSPC chart

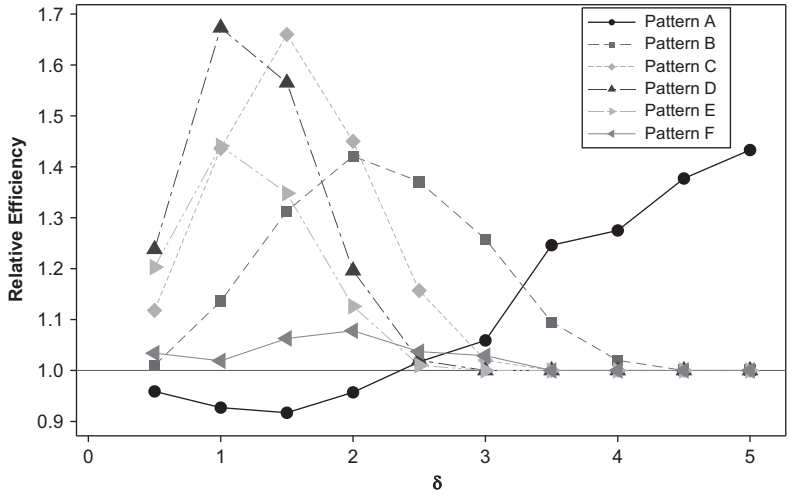
In general, design of the FL-MSPC chart involves three parameters, λ_1, λ_2 and h. Once the parameters λ_1 and λ_2 are chosen, h is determined by the pre-specified value of ARL_0. In order to gain insights intothe joint effects of the two parameters λ_1 and λ_2 on the run length performance of the FL-MSPC chart, we simulate the clusters in Figure 4 and evaluate the ARL values of the corresponding FL-MSPC charts for detecting spatial clusters on a 2-D map.

To thoroughly study the performance of the proposed chart under different parameter settings, we here set λ_1 to four different levels, and λ_2 to five different levels. Thus, 20 parameter combinations, as shown in Table 1, are generated in total. For each parameter combination, the corresponding FL-MSPC chart is implemented to monitor a simulated process with 6 different shift clusters with 10 shift sizes; their out-of-control ARL performance is reported in Table 2. All control limits are tuned to have an equal in-control ARL approximately 200.

From Table 2, it can be seen that imposing a small L_1-penalty when $\lambda_2 = 0.0$ always helps detecting spatial clusters although the improvement over the conventional T^2 chart is not significant. As shown in Figure 5, the relative efficiency of the FL-MSPC chart may deteriorate for detecting medium to large clusters with small magnitude when λ_1 is too big. When fixing the value of λ_1, increasing the value of λ_2 often helps detecting large clusters. For example, for detecting a cluster of



(a) $\lambda_1 = 0.2$ and $\lambda_2 = 0.0$



(b) $\lambda_1 = 0.0$ and $\lambda_2 = 0.2$

Figure 5. RE of FL-MSPC charts with only one tuning parameter.

Table 1. Parameter combinations of the FL-MSPC chart to be studied.

Case number	λ_1	λ_2	Case number	λ_1	λ_2
1	0	0	11	0.1	0
2	0	0.1	12	0.1	0.1
3	0	0.2	13	0.1	0.2
4	0	0.3	14	0.1	0.3
5	0	0.4	15	0.1	0.4
6	0.05	0	16	0.2	0
7	0.05	0.1	17	0.2	0.1
8	0.05	0.2	18	0.2	0.2
9	0.05	0.3	19	0.2	0.3
10	0.05	0.4	20	0.2	0.4

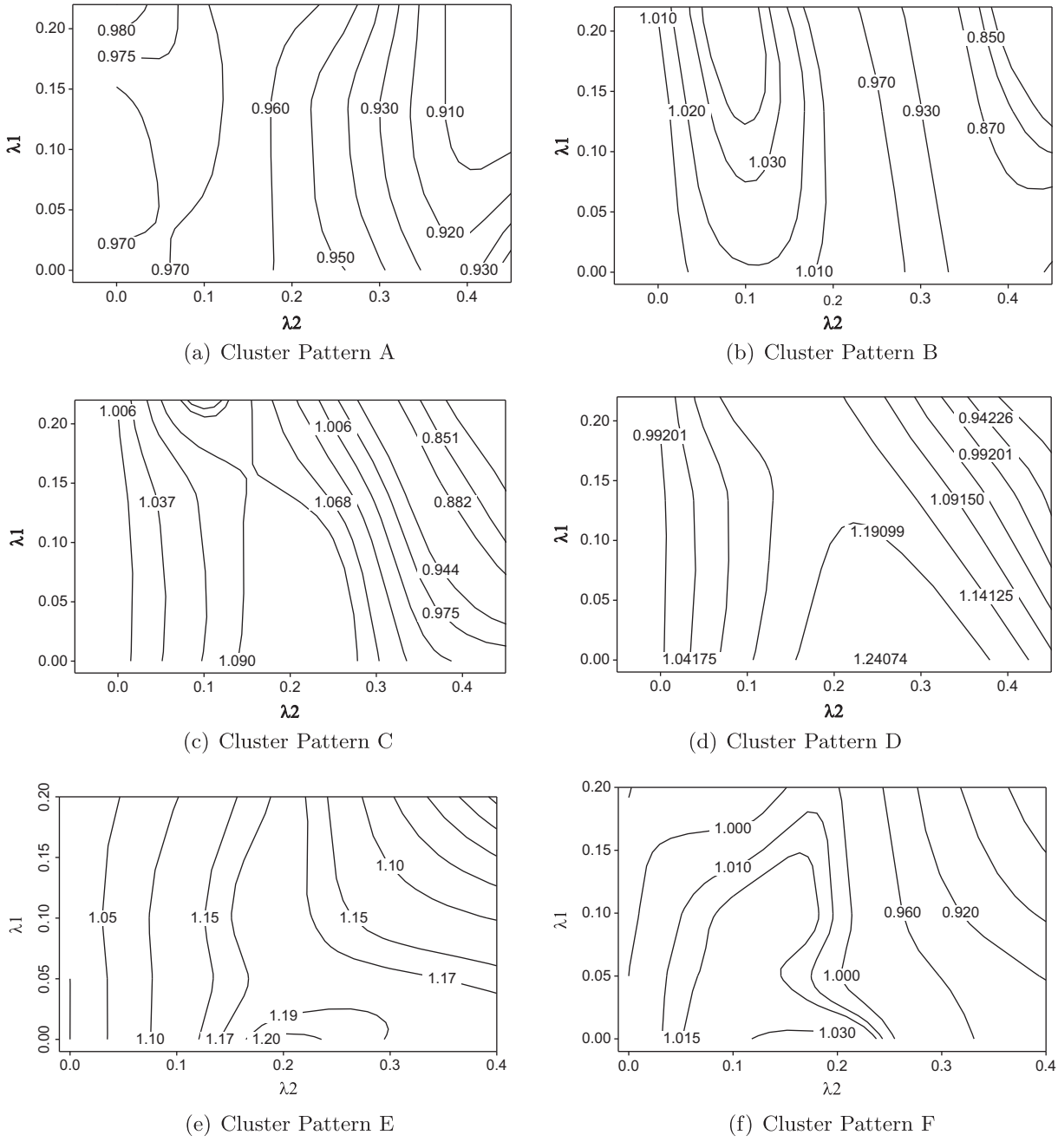
pattern C with $\delta = 1$, the FL-MSPC chart with $\lambda_1 = 0.1$ takes 31.9 steps on average if $\lambda_2 = 0.0$, which is significantly longer than the FL-MSPC chart with $\lambda_2 = 0.2$, which only takes 22.8 steps on average. However, too much penalty using λ_2 is too

Table 2. ARL performance of the FL-MSPC chart.

λ_1	0.0					0.05					0.1					0.2				
λ_2	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
Cluster Pattern A																				
$\delta = 0.5$	188	188	191	195	198	189	188	192	191	197	189	187	192	195	201	187	187	193	195	197
1.0	158	160	169	179	189	158	160	171	177	188	157	160	170	180	194	156	160	170	181	191
1.5	118	124	133	152	173	118	124	134	150	173	118	124	134	153	178	116	123	131	152	176
2.0	80.9	87.4	91.1	110	142	80.8	86.4	91.1	109	142	79.9	85.8	89.8	109	143	78.3	85.0	87.8	108	142
2.5	51.9	52.9	53.7	63.2	92.2	51.6	52.6	52.3	61.4	88.2	50.9	51.9	51.7	61.4	83.1	49.5	50.9	49.1	56.7	79.7
3.0	31.9	31.4	31.4	33.8	44.4	31.6	31.2	30.2	31.4	41.9	31.2	30.5	29.6	30.1	39.3	30.2	29.7	28.1	26.5	35.7
3.5	19.5	19.3	16.8	16.3	21.0	19.2	18.6	16.5	15.7	19.2	18.8	18.0	15.9	14.9	17.8	18.0	17.4	14.4	13.2	15.3
4.0	12.0	11.2	9.89	9.09	9.95	11.8	11.1	9.55	8.55	9.15	11.5	10.8	9.23	8.07	8.53	11.1	10.4	8.45	7.04	7.20
4.5	7.46	7.07	5.97	5.13	5.31	7.29	6.93	5.74	4.81	4.87	7.11	6.71	5.50	4.56	4.45	6.73	6.33	4.88	3.88	3.83
5.0	4.71	4.51	3.63	3.07	3.07	4.61	4.35	3.50	2.81	2.80	4.52	4.19	3.35	2.67	2.66	4.29	3.92	3.00	2.41	2.31
Cluster Pattern B																				
$\delta = 0.5$	149	147	148	156	165	149	147	147	157	168	149	145	145	161	170	148	147	147	164	179
1.0	65.4	63.3	59.4	68.7	88.0	65.6	63.3	59.3	69.5	92.2	65.2	62.5	59.6	72.8	97.0	64.5	62.3	60.1	78.0	117
1.5	22.2	20.7	18.1	19.4	29.2	22.2	20.5	17.9	19.3	29.7	22.0	20.1	17.8	19.6	31.2	21.7	19.8	17.3	20.7	36.9
2.0	7.49	6.94	5.80	5.68	7.94	7.44	6.89	5.67	5.53	7.73	7.33	6.78	5.55	5.45	7.86	7.18	6.66	5.26	5.39	8.35
2.5	2.99	2.75	2.37	2.15	2.63	2.96	2.70	2.30	2.10	2.52	2.93	2.65	2.23	2.07	2.46	2.87	2.60	2.15	2.01	2.51
3.0	1.59	1.50	1.32	1.25	1.35	1.58	1.48	1.30	1.24	1.33	1.57	1.46	1.29	1.23	1.32	1.54	1.43	1.25	1.21	1.31
3.5	1.14	1.10	1.05	1.03	1.04	1.14	1.09	1.04	1.03	1.04	1.13	1.08	1.04	1.03	1.03	1.12	1.07	1.03	1.02	1.03
4.0	1.03	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cluster Pattern C																				
$\delta = 0.5$	118	111	104	110	117	118	111	104	112	121	118	110	104	116	129	117	112	104	126	148
1.0	31.9	28.2	22.7	22.3	28.5	32.0	28.1	22.9	22.3	29.5	31.9	27.9	22.8	23.1	32.0	31.5	27.5	22.6	24.9	39.3
1.5	7.33	6.11	4.80	4.47	5.39	7.32	6.07	4.74	4.36	5.40	7.27	6.02	4.69	4.41	5.48	7.16	5.92	4.57	4.43	6.14
2.0	2.29	2.06	1.71	1.55	1.73	2.28	2.04	1.69	1.53	1.68	2.27	2.02	1.67	1.53	1.66	2.23	1.97	1.64	1.52	1.73
2.5	1.22	1.13	1.08	1.03	1.04	1.22	1.12	1.07	1.03	1.04	1.21	1.12	1.07	1.03	1.04	1.21	1.11	1.06	1.03	1.05
3.0	1.02	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00	1.02	1.01	1.00	1.00	1.00	1.02	1.00	1.00	1.00	1.00
3.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2. (*Continued*).

λ_1	0.0					0.05					0.1					0.2				
λ_2	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
Cluster Pattern D																				
$\delta = 0.5$	95.2	83.9	75.8	77.1	79.5	95.6	83.9	76.7	75.6	80.9	95.1	83.9	78.0	80.3	86.7	94.4	85.9	77.6	89.4	111
1.0	17.4	14.5	10.8	9.60	11.1	17.4	14.1	10.8	9.6	11.3	17.3	14.0	10.7	9.71	11.8	17.1	13.8	10.5	10.4	15.5
1.5	3.40	2.88	2.32	2.04	2.21	3.40	2.86	2.31	2.01	2.20	3.38	2.82	2.28	2.00	2.20	3.34	2.80	2.21	2.04	2.46
2.0	1.32	1.22	1.12	1.07	1.08	1.32	1.21	1.11	1.06	1.08	1.32	1.21	1.10	1.06	1.08	1.31	1.19	1.10	1.06	1.10
2.5	1.02	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	1.00
3.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cluster Pattern E																				
$\delta = 0.5$	94.6	84	79	80	79.6	94.6	84.3	80.2	80.4	81.6	93.9	84.1	80.1	83.4	86.3	94.1	86.1	80.6	89.6	106.9
1.0	16.2	13	11	12	13.8	16.1	13.5	11.3	11.6	14.0	16.1	13.5	11.2	11.8	14.6	16.1	13.6	11.2	13.1	19.0
1.5	3.02	2.59	2.24	2.31	2.88	3.01	2.58	2.24	2.32	2.91	2.99	2.56	2.23	2.30	2.97	2.97	2.56	2.19	2.37	3.43
2.0	1.25	1.15	1.11	1.11	1.17	1.24	1.15	1.11	1.10	1.17	1.24	1.15	1.10	1.10	1.18	1.23	1.15	1.10	1.11	1.22
2.5	1.01	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cluster Pattern F																				
$\delta = 0.5$	127	124	123	131	133	127	125	127	133	139	128	125	126	137	145	127	129	127	142	157
1.0	38.1	36.9	37.4	42.0	45.7	38.1	37.2	37.6	42.4	48.8	37.9	36.9	37.9	44.0	52.6	37.4	37.2	38.3	48.7	66.0
1.5	10.0	9.4	9.4	10.8	13.3	9.89	9.44	9.31	10.7	13.5	9.84	9.40	9.22	10.8	13.9	9.69	9.39	9.11	11.1	17.8
2.0	2.91	2.7	2.7	3.1	4.05	2.87	2.73	2.69	3.06	4.06	2.86	2.70	2.66	3.02	4.16	2.81	2.68	2.56	3.10	4.69
2.5	1.40	1.4	1.4	1.4	1.72	1.39	1.35	1.34	1.39	1.70	1.38	1.34	1.33	1.38	1.71	1.37	1.34	1.30	1.36	1.79
3.0	1.06	1.1	1.0	1.1	1.10	1.06	1.05	1.03	1.05	1.10	1.06	1.05	1.03	1.05	1.10	1.05	1.04	1.03	1.04	1.12
3.5	1.00	1.0	1.0	1.0	1.01	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.01
4.0	1.00	1.0	1.0	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.0	1.0	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.0	1.0	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00



(a) Cluster Pattern A (b) Cluster Pattern B

(c) Cluster Pattern C (d) Cluster Pattern D

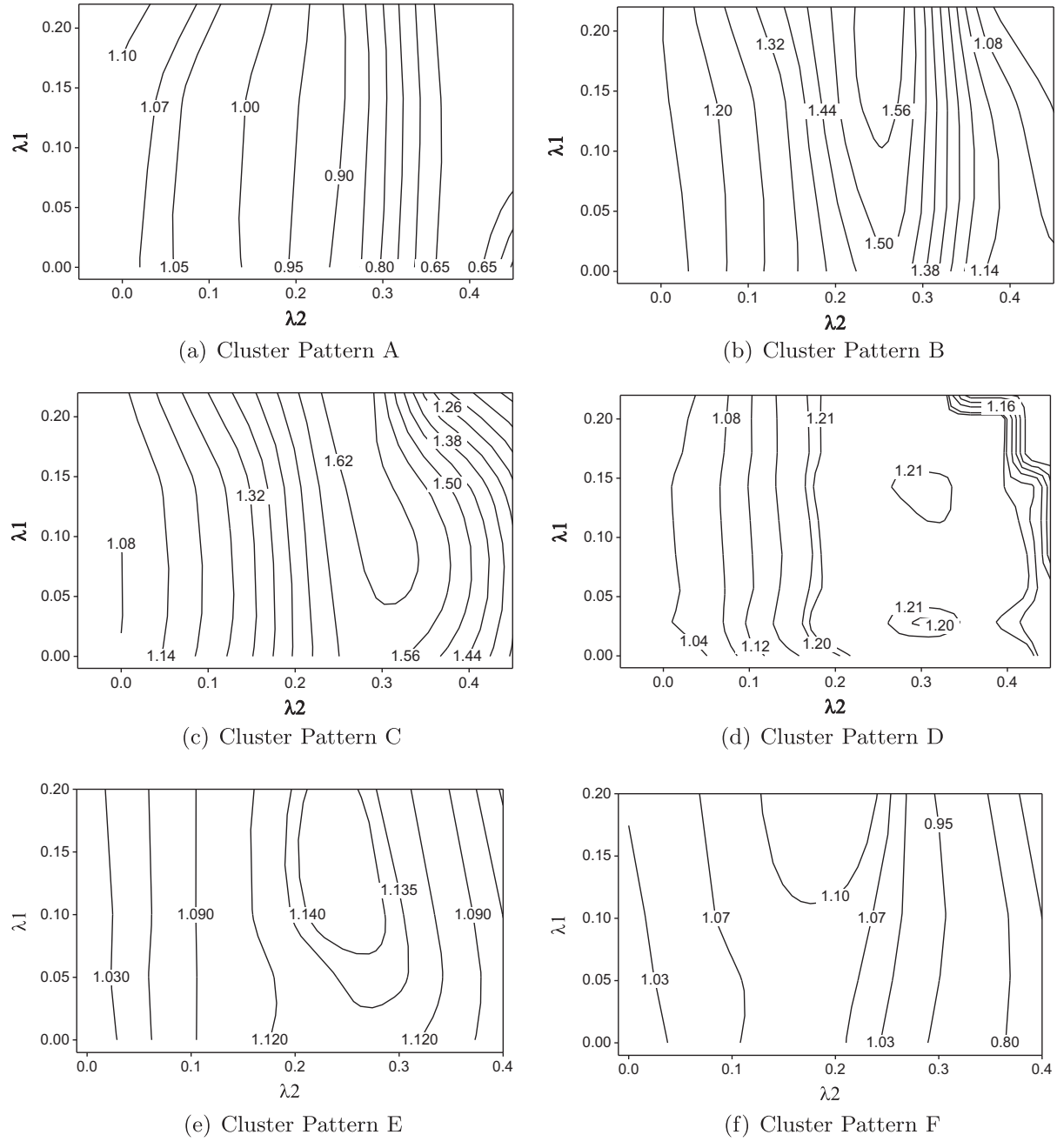
(e) Cluster Pattern E (f) Cluster Pattern F

Figure 6. RE contour plots of relative efficiency for FL-MSPC charts when $\delta = 0.5$.

conservative in cluster estimation, which in turn hurts the detection of the cluster. For the same value of λ_1, if $\lambda_2 = 0.4$, the corresponding FL-MSPC chart takes 32 steps on average to detect the cluster.

It is interesting to note that, for detecting mean shifts on a single site, the introduction of $\lambda_2 > 0$ always deteriorates the detection capability of the FL-MSPC chart when the shift magnitude is small or moderate, and improves the capability when the shift magnitude is large, regardless the value of λ_1. For example, the ARL value of the FL-MSPC chart with $\lambda_1 = 0.2$ increases from 156 to 191 for detecting a single-site mean shift of magnitude $\delta = 1$ when λ_2 increases from 0.0 to 0.4. On the other hand, the ARL of the same charts decreases from 11.1 to 7.20 for $\delta = 4$. This illustrates that the penalty imposed by λ_2 can also helps detecting large magnitudes of single-site mean shifts, although it mainly improves the detection of clusters.

Based on the above discussion of the fused LASSO algorithm for cluster identification, we know that the choice of λ_1 and λ_2 strongly depends on map configuration, spatial cluster size and shift magnitude. In general, there is no selection of the parameters that is uniformly better than other choices. To facilitate the design of FL-MSPC charts, Figures 6 and 7 show



(a) Cluster Pattern A (b) Cluster Pattern B

(c) Cluster Pattern C (d) Cluster Pattern D

(e) Cluster Pattern E (f) Cluster Pattern F

Figure 7. Contour plots of relative efficiency for FL-MSPC charts when $\delta = 2$.

contour plots of the RE of FL-MSPC charts for detecting the six cluster patterns with $\delta = 0.5$ and 2.0, respectively. If one only concerns cluster detection, we should refer to plots (b)–(f) in Figures 6 and 7. For detecting clusters with small shift magnitudes (e.g. $\delta = 0.5$), the optimal choice of (λ_1, λ_2) changes from somewhere around (0.1, 0.2) to somewhere around (0.0, 0.2) when the cluster size increases from small to large. Similarly, for detecting clusters with moderate shift magnitudes (e.g. $\delta = 2$), the optimal choice of (λ_1, λ_2) remains around (0.2, 0.3) when the cluster size increases from small to large, and (0.1, 0.2) for the complex shift patterns. For detecting clusters with moderate shift magnitudes ($\delta \geq 3$), the FL-MSPC chart is not sensitive to parameter choices for large clusters since all cluster shifts can often be detected right after the change happens. Therefore, if a simple guideline for parameter choice is preferred, a choice of (0.1, 0.2) for detecting small clusters and (0.05, 0.3) or (0.1, 0.2) may work well for detecting large clusters in practice. Note that these choices do not intend to optimise the run length performance for detecting any particular magnitude of shifts. In the following, we should compare the performance of these two FL-MSPC charts with other alternative control charts for detecting spatial clusters.

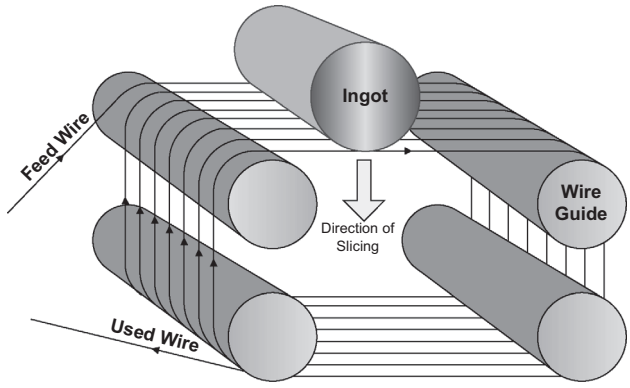


Figure 8. Wire slicing process.

6. Comparison with other alternative control charts

The FL-MSPC chart is designed to detect clusters in a 2-D map. We now compare the FL-MSPC chart with several alternative multivariate control charts. The list of competing charts should cover both publicly popular methods seen in the SPC literature and practical methods really used by field engineers.

As discussed before, if we treat the 2-D map as a high-dimensional vector, we can implement the Hotelling's T^2 chart. Hotelling's T^2 chart is a popular chart for general purpose multivariate process monitoring. Therefore, we choose Hotelling's T^2 chart as the first alternative.

The second alternative is the common practice for STIR monitoring. In practice, an engineer only picks up the largest STIR value one sees from a map and monitor the extreme value. This is essentially a Shewhart chart that monitors the maximum STIR value of each wafer (denoted as the Max chart here). It can also be considered as a variable selection-based control chart since it selects one site each time. As far as we know, the integrated circuit industry ignores rich 2-D information but looks at the maximum of all sites only for quality inspection. Obviously, this treatment results in significant loss of useful information that have already been obtained from wafer surface.

The third chart for comparison is the variable selection-based chart (VS-MSPC) proposed by Wang and Jiang (2009). The VS-MSPC chart is established based on the idea of achieving dimension reduction via variable selection. However, the VS-MSPC chart does not consider the spatial structure that a 2-D data sample possesses, which may potentially adversely affect its performance. There is one critical parameter, M, in the VS-MSPC chart, that is, the number of suspected sites. The VS-MSPC chart only keeps the exact number of suspected sites during variable selection. Wang and Jiang (2009) did not give an optimal choice for parameter M. The authors tested $M = 1 \sim 3$ for processes with 50 and 100 variables, and suggested that the VS-MSPC chart is robust to parameter misspecification; the performance of the VS-MSPC chart is visually indistinguishable when underestimate or overestimate the true number of shifted variables. Therefore, a small value is usually suggested to benefit from the enhanced power of reducing dimension. In the following simulation, for demonstration, M is set to five.

For comparison purposes, we select three FL-MSPC charts , (0.2, 0.0), (0.1, 0.2) and (0.05, 0.3). The first one is a counterpart of the VS-MSPC chart, which uses L_1- instead of L_0-penalty for variable selection. The other two FL-MSPC chart with nonzero value of λ_2 follow from the design in the last section and aim to detect small and large clusters, respectively. The ARL comparison results are shown in Table 3. It is easy to see that no chart is uniformly the best among all alternatives for detecting any clusters of mean shifts. The industry standard method, the Max chart performs the best only for detecting mean shifts of a single site. This is because the Max chart selects the most outstanding site each time and thus greatly reduces the number of dimension. However, when there is clusters, the Max chart ignores useful information in other sites and always performs the worst among all alternatives. That is, it is not efficient for cluster identification.

Comparing to the T^2 chart, the VS-MSPC chart performs better for mean shifts of moderate to large magnitudes in small clusters, but worse when the mean shifts are small and/or in large clusters. Wang and Jiang (2009) have argued that the VS-MSPC chart performs worse than the T^2 chart due to the randomness of variable selection algorithms when small shifts occur. For detecting large clusters, since the VS-MSPC chart always keeps a limited number of suspected sites (e.g. 5 in our example), when the true number of shifted sites are much larger than the designated one, its performance is often hurt.

As a special case, comparing the FL-MSPC chart with parameters (0.2, 0.0) and the VS-MSPC chart, both charts utilise variable selection techniques to reduce dimensions. The latter uses L_0-penalty with the specified number of suspected variables, while the former uses L_1-penalty with the penalty coefficient λ_1. The larger value λ_1 is, less variables will be

Table 3. Comparison of control charts for detecting spatial clusters.

Method	T^2	Max	VS-MSPC	FL-MSPC		
λ_1	–	–	–	0.2	0.1	0.05
λ_2	–	–	–	0.0	0.2	0.3
			Cluster Pattern A			
$\delta = 0.5$	188	*186*	192	187	192	191
1.0	155	*146*	156	156	170	177
1.5	118	*72.4*	97.6	116	134	150
2.0	80.9	*27.4*	48.4	78.3	89.8	109
2.5	51.9	*11.9*	23.2	49.5	51.7	61.4
3.0	31.9	*5.25*	10.6	30.2	29.6	31.4
3.5	19.5	*2.85*	5.16	18.0	15.9	15.7
4.0	12.0	*1.80*	3.09	11.1	9.23	8.55
4.5	7.46	*1.36*	2.05	6.73	5.50	4.81
5.0	4.71	*1.16*	1.51	4.29	3.35	2.81
			Cluster Pattern B			
$\delta = 0.5$	149	158	163	148	*145*	157
1.0	65.4	69.1	72.1	64.5	*59.6*	69.5
1.5	22.2	21.4	19.0	21.7	*17.8*	19.3
2.0	7.49	7.14	*5.44*	7.18	5.55	5.53
2.5	2.99	2.85	*1.98*	2.87	2.23	2.10
3.0	1.59	1.53	*1.20*	1.54	1.29	1.24
3.5	1.14	1.13	*1.03*	1.12	*1.04*	*1.03*
4.0	1.03	1.02	1.00	1.02	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00
			Cluster Pattern C			
$\delta = 0.5$	118	132	136	117	*104*	112
1.0	31.9	44.6	40.9	31.5	22.8	*22.3*
1.5	7.33	12.7	8.30	7.16	4.69	*4.36*
2.0	2.29	4.29	2.26	2.23	1.67	*1.53*
2.5	1.22	1.90	1.17	1.21	1.07	*1.03*
3.0	1.02	1.16	1.01	1.02	1.00	1.00
3.5	1.00	1.02	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00
			Cluster Pattern D			
$\delta = 0.5$	95.2	113	115	94.4	78.0	*75.6*
1.0	17.4	32.7	27.0	17.1	10.7	*9.60*
1.5	3.40	8.80	4.45	3.34	2.28	*2.01*
2.0	1.32	3.08	1.55	1.31	1.10	*1.06*
2.5	1.02	1.49	1.04	1.02	1.00	1.00
3.0	1.00	1.06	1.00	1.00	1.00	1.00
3.5	1.00	1.00	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00

Table 3. (*Continued*).

Method	T^2	Max	VS-MSPC	FL-MSPC		
λ_1	–	–	–	0.2	0.1	0.05
λ_2	–	–	–	0.0	0.2	0.3
			Cluster Pattern E			
$\delta = 0.5$	94.6	113	110	94.1	*80.1*	80.4
1.0	16.2	31.6	24.6	16.1	*11.2*	11.6
1.5	3.02	8.09	4.03	2.97	*2.23*	2.32
2.0	1.25	2.74	1.44	1.23	*1.10*	*1.10*
2.5	1.01	1.38	1.03	1.00	1.00	1.00
3.0	1.00	1.04	1.00	1.00	1.00	1.00
3.5	1.00	1.00	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00
			Cluster Pattern F			
$\delta = 0.5$	127	140	142	127	*126*	133
1.0	38.1	49.4	47.1	*37.4*	37.9	42.4
1.5	9.95	14.4	9.98	9.69	*9.22*	10.7
2.0	2.91	4.51	2.68	2.81	*2.66*	3.06
2.5	1.40	1.91	*1.24*	1.37	1.33	1.39
3.0	1.06	1.19	*1.02*	1.05	1.03	1.05
3.5	1.00	1.03	1.00	1.00	1.00	1.00
4.0	1.00	1.00	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00	1.00	1.00

selected as suspicious variables. When $\lambda_1 \to 0$, the FL-MSPC chart reduces to the T^2 chart. In our comparison, the VS-MSPC chart only outperforms the FL-MSPC chart with $\lambda_1 = 0.2$ and $\lambda_2 = 0$ for shifts of moderate to large magnitudes in Pattern A.

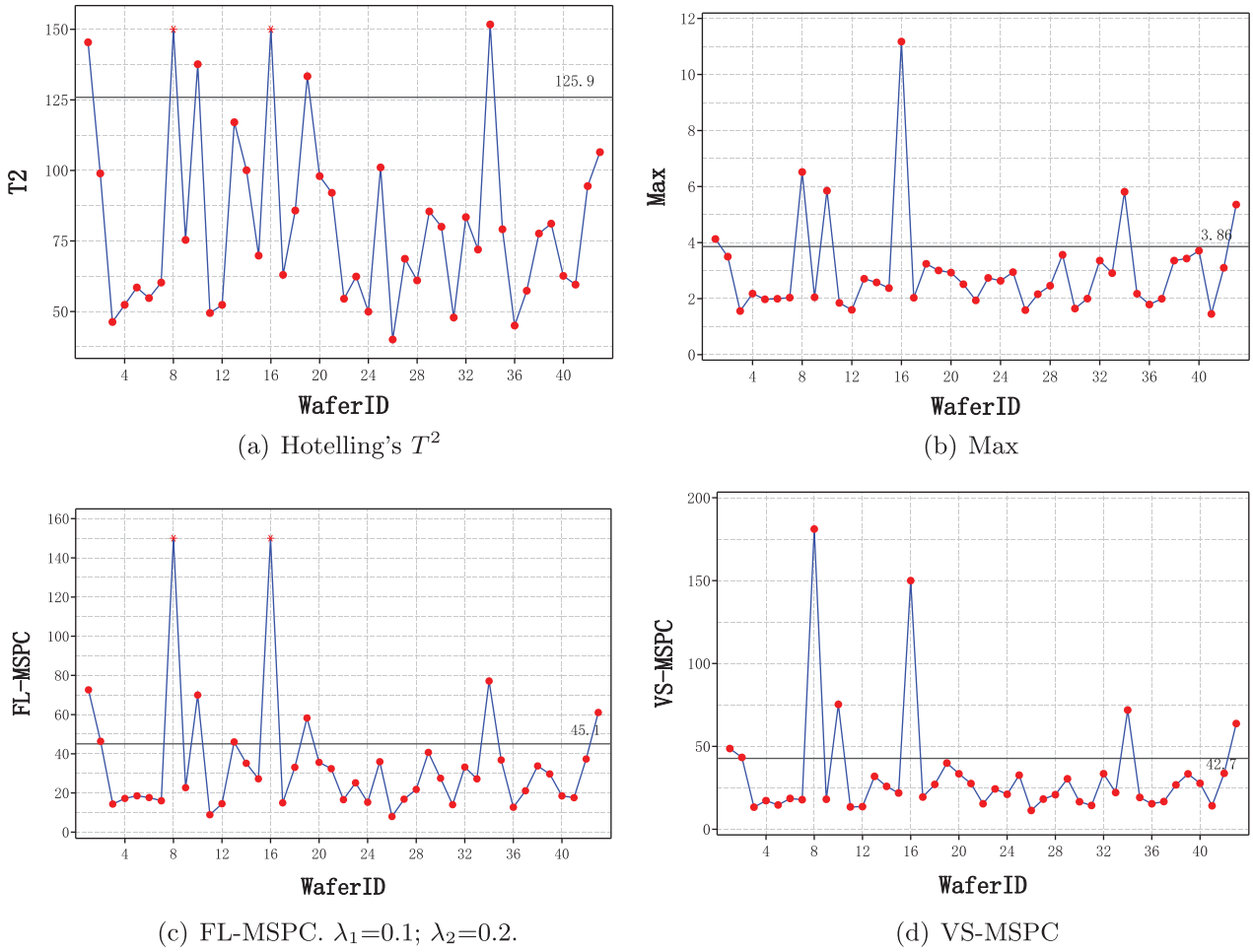
As expected, the FL-MSPC chart usually performs the best for detecting clusters of shifts. The FL-MSPC chart with parameters (0.1, 0.2) gives the lost ARL for Pattern B, and the FL-MSPC chart with parameters (0.05, 0.3) performs almost the best for detecting shift Patterns C and D. For the more complex shift patterns E and F, the FL-MSPC chart still performs the best among all.

In summary, if the shift pattern contains only one site, which means the number of shifted variables is very small in the large-dimensional 2-D map, the Max chart that simply monitors the maximum value among all sites is efficient in detecting such shifts. While if shifts are likely to occur in clusters, the proposed FL-MSPC chart is more efficient in detecting such clusters in a 2-D map. The added fusion and LASSO penalties can successfully capture the shift patterns, thus help in reducing dimensionality which in turn improves the detection performance of the FL-MSPC chart. In the next section, we will apply the proposed method to a real industry example to demonstrate its effectiveness.

7. An real example in semiconductor manufacturing

In this section, we apply the proposed chart to the wafer example. One critical step that affects the STIR values of a wafer is the wire slicing stage in wafer fabrication. A schematic illustration of this process is shown in Figure 8. Cutting wire is winded around four wire guides and fed in a forward/backward way, thus functioning as a saw; a silicon ingot moves downwards and is cut into thin pieces by the wire saw.

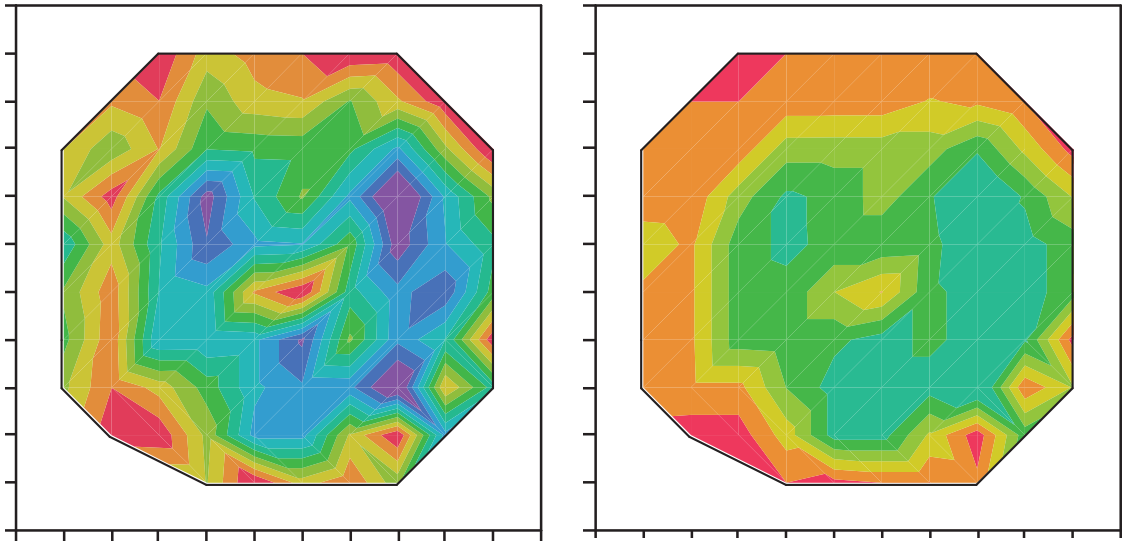
In the slicing process, wafer quality is very sensitive to process conditions. For example, slurry used for slicing is critical to surface roughness. Slurry is usually blended first and moved to the slicing machine for use. If the slurry is not properly blended, sustained mean shifts in wafer quality is expected to be seen. As another example, tension force is a critical factor that affect wafer flatness. Tension force is maintained by a tension unit. If the unit shifts or fails, wafer flatness is expected to be affected, and this shift will remain for several batches until the tension unit is identified and properly tuned or replaced. Therefore, control chart should be applied to monitor slicing wafers and alarm potential process shifts. In fact, a Max chart



(a) Hotelling's T^2 (b) Max

(c) FL-MSPC. λ_1=0.1; λ_2=0.2. (d) VS-MSPC

Figure 9. Control charts for a cut of 43 wafers.



(a) Original STIR map (b) STIR map after applying fused LASSO. λ_1=0.1; λ_2=0.2.

Figure 10. Heat map of STIR values before and after applying the fused LASSO algorithm.

has been equipped by engineers in practice to monitor the slicing process. As previous studied, the Max chart is insensitive to clustered shifts. In the following, we applied different control charts to monitor real samples taken from the slicing process, and demonstrate the usage of the proposed chart.

In the wafer testing site, 43 wafers are measured. A single cut of 43 wafers provides 43 2-D STIR maps. Since at the time of our data collection, process parameters have not been optimally adjusted and the process is not in control. Therefore, we expect all wafers to be out of control. We applied the proposed and competing control charts to monitor these samples and identified suspicious clusters, as shown in Figure 9. Note that, the applications of these control charts are essentially the Shewhart-type, which can be considered as a tool for testing one-at-a-time rather than sequential testing. It is found that wafer ID #8 and #16 on (a) and (c) have extremely large values and thus have been rescaled. Not surprisingly, we found quite a few out-of-control signals in most charts. Specifically, the FL-MSPC chart identifies 9 out-of-control wafers, the VS-MSPC chart identifies 7, while the T^2 and Max chart identify 6 samples to be out-of-control, respectively.

We took a close look at wafer ID #8 that is declared to be out-of-control by all charts, as shown in Figure 9. Figure 10 further shows the heat map of the original dataset and mean estimates obtained from the fused LASSO algorithm. It is evident that the original STIR map shows scattered clusters. After applying the fused LASSO algorithm, some adjacent sites are combined and form large clusters. Further industrial experience reveals that these defective wafers are caused by uneven processing conditions near-edge areas.

The numerical results show that the fused LASSO algorithm is capable of capturing the shift pattern by removing noise and making the shift signals stand out. The original shift pattern, in clusters or other complex shapes, are maintained by the algorithm. This favourable feature of the fused LASSO provides important information to fault diagnosis in practice. As root cause diagnosis is always followed when an out-of-control signal is triggered, it would be beneficial to provide clues regarding the shape, location, size and other aspects of the patterns to guide further investigation of the failures and save cost and time of such operations. The above example shows that the FL-MSPC chart has the merit of providing needed diagnostic information, which is one important advantageous of this proposed chart.

8. Conclusions

This paper proposes a new SPC method for monitoring product surface. Due to special data structures, shifts on product surfaces tend to form spatial clusters. The proposed scheme first uses a variable selection algorithm to estimate shift sites. A fusion penalty is added to help identifying spatial clusters and a L_1-penalty is added to reduce noises. A directional chart is then applied to detect out-of-control clusters. Our simulation studies show that the added penalties are beneficial in identifying clusters, enhancing control chart performance, and providing diagnostic information for root cause analysis. An industry example from the wafer fabrication process is investigated to demonstrate the usefulness of the proposed method; the use of the proposed chart can help achieve not only a quick detection of process faults, but also an accurate identification of root causes.

In this paper, we have successfully adapted the fused LASSO algorithm for spatial shift estimation, and studied its design issues. The performance of a likelihood ratio-based control chart has been studied; practical guidelines for control chart selection and design have been presented. The monitoring of 2-D data is a challenging work that needs more research efforts. For example, this paper only considers one test at a time using the current map information, and demonstrates the usefulness of a variable-selection method for process monitoring and diagnosis in 2-D maps. Similar to Zou and Qiu (2009) and Jiang, Wang, and Tsung (2012), it is not difficult to extend our framework to using multivariate EWMA statistics by cumulating historical information. This is expected to improve the SPC detection performance when the smoothing parameter is properly chosen. In addition, shifts in 2-D data maps possesse unique spatial structures; measures obtained from adjacent sites are usually correlated. Integrating such information with control chart design should further benefit its performance.

Acknowledgements

The authors are grateful to the Editor-in-Chief, Prof. Alexandre Dolgui, for his insightful comments and three anonymous referees for their valuable comments, which have helped us improve this work greatly.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Kaibo Wang's work was supported by the National Natural Science Foundation of China (NSFC) [grant number 71471096]. Wei Jiang's work was partially supported by NSFC [grant number 71172131], [grant number 71325003], [grant number 71531010]; the Program of Shanghai Subject Chief Scientist [grant number 15XD1502000]. Bo Li's work was partly supported by the Beijing Higher Education Young Elite Teacher Project under Grant YETP0135 and NSFC [grant number 71490723], [grant number 71432004], [grant number 71272029].

References

Albin, S., and D. Friedman. 1989. "The Impact of Clustered Defect Distributions in IC Fabrication." *Management Science* 35 (9): 1066–1078.

Bao, L., K. Wang, and R. Jin. 2014. "A Hierarchical Model for Characterizing Spatial Wafer Variations." *International Journal of Production Research* 52 (6): 1827–1842.

Cheng, L., P. Gupta, C. J. Spanos, K. Qian, and L. He. 2011. "Physically Justifiable Die-level Modeling of Spatial Variation in View of Systematic Across Wafer Variability." *IEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 30 (3): 388–401.

Chien, C. F., S. Hsu, and Y. J. Chen. 2013. "A System for Online Detection and Classification of Wafer Bin Map Defect Patterns for Manufacturing Intelligence." *International Journal of Production Research* 51 (8): 2324–2338.

Colosimo, B. M., Q. Semeraro, and M. Pacella. 2008. "Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles." *Journal of Quality Technology* 40 (1): 1–18.

Friedman, J., T. Hastie, H. Hoefling, and R. Tibshirani. 2007. "Pathwise Coordinate Optimization." *Annals of Applied Statistics* 1 (2): 302–332.

Hansen, M. H., V. N. Nair, and D. J. Friedman. 1997. "Monitoring Wafer Map Data from Integrated Circuit Fabrication Processes for Spatially Clustered Defects." *Technometrics* 39 (3): 241–253.

Hoefling, H. 2010. "A Path Algorithm for the Fused LASSO Signal Approximator." *Journal of Computational and Graphical Statistics* 19 (4): 984–1006.

Hsu, S. C., and C. F. Chien. 2007. "Hybrid Data Mining Approach for Pattern Extraction from Wafer Bin Map to Improve Yield in Semiconductor Manufacturing." *International Journal of Production Economics* 107 (1): 88–103.

Hwang, J. Y., and W. Kuo. 2007. "Model-Based Clustering for Integrated Circuit Yield Enhancement." *European Journal of Operational Research* 178: 143–153.

Jeong, M. K., J. C. Lu, and N. Wang. 2006. "WaveletBased SPC Procedure for Complicated Functional Data." *International Journal of Production Research* 44 (4): 729–744.

Jeong, Y., S. J. Kim, and M. K. Jeong. 2008. "Automatic Identification of Defect Patterns in Semiconductor Wafer Maps Using Spatial Correlogram and Dynamic Time Warping." *IEEE Transaction on Semiconductor Manufacturing* 21 (4): 625–637.

Jiang, W., and K.-L. Tsui. 2008. "A Theoretical Framework and Efficiency Study of Multivariate Control Charts." *IIE Transactions on Quality and Reliability* 40 (7): 650–663.

Jiang, W., K. Wang, and F. Tsung. 2012. "A Variable-Selection-based Multivariate EWMA Chart for Process Monitoring and Diagnosis." *Journal of Quality Technology* 44 (3): 209–230.

Lin, J., and K. Wang. 2011. "Online Parameter Estimation and Run-to-Run Process Adjustment Using Categorical Observations." *International Journal of Production Research* 49 (13): 4103–4117.

Lin, B., K. Wang, and A. B. Yeh. 2013. "Monitoring the Covariance Matrix via Penalized Likelihood Estimation." *IIE Transactions* 45: 132–146.

Liu, S. F., F. L. Chen, Y. Y. Shi, S. M. Yu, and C. S. Chang. 2008. "Wavelet Transform Based Wafer Defect Map Pattern Recognition System in Semiconductor Manufacturing." Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, 19–21.

Montgomery, D. 2009. *Introduction to Statistical Quality Control*. 6th ed. New York: Wiley.

Reda, S., and S. R. Nassif. 2009. "Analyzing the impact of Process Variations on Parametric Measurements: Novel Models and Applications." Proceedings of the Conference on Design, Automation and Test in Europe, Nice, 375–380.

Rudin, L. I., S. Osher, and E. Fatemi. 1992. "Nonlinear Total Variation Based Noise Removal Algorithms." *Physica D: Nonlinear Phenomena* 60: 259–268.

Stine, B. E., D. S. Boning, and J. E. Chung. 1997. "Analysis and Decomposition of Spatial Variation in Integrated Circuit Processes and Devices." *IEEE Transactions on Semiconductor Manufacturing* 10: 24–41.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the LASSO." *Journal of the Royal Statistical Society Series B-Methodological* 58 (1): 267–288.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight. 2005. "Sparsity and Smoothness via the Fused LASSO." *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 67 (1): 91–108.

Wang, K., and W. Jiang. 2009. "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection." *Journal of Quality Technology* 41 (3): 247–258.

Wang, K., and F. Tsung. 2008. "An Adaptive T^2 Chart for Monitoring Dynamic Systems." *Journal of Quality Technology* 40: 109–123.

Wang, X., S. Wu, and K. Wang. 2015. "A Run-to-Run Profile Control Algorithm for Improving the Flatness of Nano-scale Products." *IEEE Transactions on Automation Science and Engineering* 12 (1): 192–203.

Yuan, T., and W. Kuo. 2008. "Spatial Defect Pattern Recognition on Semiconductor Wafers Using Model-based Clustering and Bayesian Inference." *European Journal of Operational Research* 190: 228–240.

Zou, C. L., and P. Qiu. 2009. "Multivariate Statistical Process Control Using LASSO." *Journal of the American Statistical Association* 104: 1586–1596.

Appendix 1. Fused LASSO algorithm

The fused LASSO algorithm consists of the following two steps (Friedman et al. 2007; Hoefling 2010) that were used in this paper,

(1) Fusion Step: Obtain $\hat{\boldsymbol{\mu}}^F$ that

$$\min_{\boldsymbol{\mu}} \left[(\boldsymbol{y}_t - \boldsymbol{\mu})'(\boldsymbol{y}_t - \boldsymbol{\mu}) + \lambda_2 \sum_{(s,t)\in E, s<t} |\mu_s - \mu_t| \right]$$

The efficient general FLSA algorithm in Hoefling (2010) starts with $\hat{\mu}_j^F = y_{tj}$, then repeatedly fuse and occasionally split neighboring locations. Eventually it produces a solution with clusters. The fusion penalty encourages cluster forming in the layout.

(2) LASSO Step

$$\tilde{\boldsymbol{\mu}} = \text{argmin}_{\boldsymbol{\mu}} \|\hat{\boldsymbol{\mu}}^F - \boldsymbol{\mu}\|^2 + \lambda_1 \sum_{j=1}^{p} |\mu_j|$$

This step is effectively a soft-shresholding operation, i.e. $\tilde{\mu}_j = sgn(\hat{\mu}_j^F)(|\hat{\mu}_j^F| - \lambda_1)_+$. The clusters obtained in the first step will be kept intact except for that those clusters with small magnitudes will be merged into a single new cluster whose magnitude is set to be 0.