# Locating Disturbances in Semiconductor Manufacturing With Stepwise Regression

Anthony T. McCray, *Student Member, IEEE*, James McNames, *Senior Member, IEEE*, and David Abercrombie

***Abstract*—The ability to locate disturbances in semiconductor manufacturing processes is critical to developing and maintaining a high yield. Analysis of variance (ANOVA), the best current practice for this problem, consists of conducting a series of hypothesis tests at each individual processing step. This approach can lead to excessive false alarms and limited sensitivity when the process contains more than one disturbance. We describe how this problem can be framed as a subset selection problem and propose two new methods based on stepwise regression. Results of over 90 000 Monte Carlo simulations suggest that these new SWR methods locate disturbances with fewer false positives and false negatives than ANOVA. This means process engineers will spend less time responding to false alarms and will be able to locate real disturbances more often.**

***Index Terms*—Analysis of variance (ANOVA), fault isolation, semiconductor manufacturing, statistical process control and monitoring, stepwise regression, variance reduction, variance source isolation.**

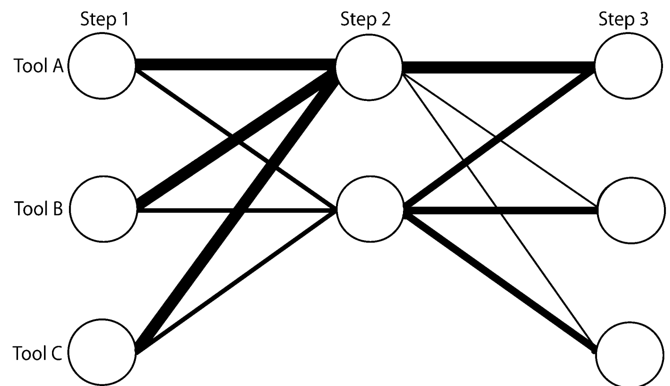



Fig. 1. Conceptual diagram of three consecutive processing steps in a semiconductor manufacturing process. Each lot is processed by a single tool at each step. The probability of making a transition between tools at adjacent steps is represented by line thickness. Thicker lines indicate higher transition probabilities.

## I. Introduction

Modern semiconductor manufacturing processes typically consist of 200 or more processing steps with multiple tools at most steps. The combined number of tools at all of the steps typically exceeds 1000. Fig. 1 shows a conceptual diagram of three consecutive steps of the process. Each lot is processed by a single tool at each processing step. Lots typically consist of 25 wafers and each wafer can have up to several thousand die. The sequence of tools at each step that process a lot, called the tool trajectory, are determined by a scheduling algorithm. Although the transition probabilities between tools are nominally uniform, in practice tools have nonuniform usage and transition probabilities, as illustrated in Fig. 1.

Tools at a single step are designed to perform the same processing. In practice, tools differ and cannot process lots identically. For processes that are well under control, the processing variability between tools at a step is minor and does not impact die quality. We define a disturbance as a tool that processes lots differently enough from other tools at the same step to impact performance of the die and yield of the lot. Process engineers face the challenge of locating steps with disturbances with little data and many possibilities. This problem applies to many different measures of the lot and die quality including electrical test (e-test) parameters, in-line metrology, and defect data.

The difficulty of this problem is compounded by several limitations of the available data. First, many measurements cannot be taken until the end of the process when the devices are complete enough to be tested and characterized electrically. Second, only lot-level statistics can be treated as statistically independent observations because all of the wafers in a lot have the same tool trajectory. Examples of lot-level statistics include yield, electrical test (e-test), die leakage, and defectivity measured between inspection points [1]. Third, the data sets often contain a small number of observations (10–100), compared to the number of steps (25–500) or step-tool combinations ($>1000$) that must be analyzed. To locate steps with disturbances, analysis techniques divide the data into subsets based on which tools at a step processed the lots. Small data sets make it difficult to estimate properties (e.g., mean and variance) of these even smaller subsets accurately. Fourth, nonuniform tool usage and transition probabilities can cause these subsets from different steps to have significant overlap. This makes them less statistically distinguishable and thereby makes it more difficult to locate disturbances.

Analysis of variance (ANOVA) is the most common method used to locate disturbances. In practice, it is applied to each of the steps consecutively. At each step, the data is grouped into subsets according to which tool processed each lot. A statistical hypothesis test is performed to determine whether the subset means are significantly different. Small data sets diminish the statistical power of ANOVA to detect disturbances. Nonuniform

Manuscript received June 15, 2004; revised March 15, 2005. This work was supported by LSI Logic.

A. T. McCray is with Sun Microsystems, Hillsboro, OR 97124 USA (e-mail: anthony.mccray@sun.com).

J. McNames is with the Integrated Circuits Design and Test Laboratory, Portland State University, Portland, OR 97207-0751 USA (e-mail: mcnames@pdx.edu).

D. Abercrombie is with Mentor Graphics Corporation, Wilsonville, OR 97070-7777 USA (e-mail: david_abercrombie@mentor.com).

Digital Object Identifier 10.1109/TSM.2005.852118

0894-6507/$20.00 © 2005 IEEE

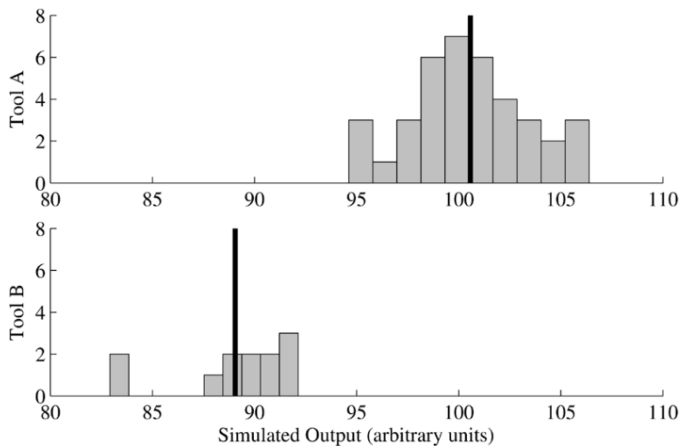



Fig. 2. These distributions represent the responses associated with tools at Step 2 of the conceptual model shown in Fig. 1. The solid black line is the average of each distribution. A disturbance has shifted the averages of these tools away from each other.

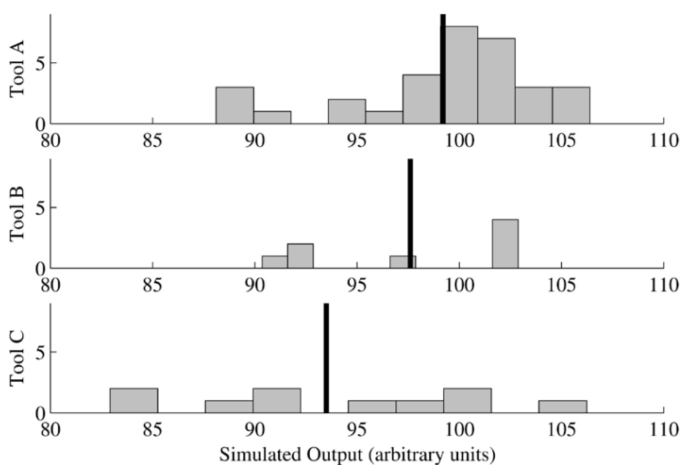



Fig. 3. Due to nonuniform mixing between Steps 2 and 3 in Fig. 1, the averages of the tools at Step 3 are statistically different. This causes ANOVA to incorrectly indicate that this step contains a disturbance (i.e., generate an FP error).

tool usage and transition probabilities can have the same effect. We quantify this problem by the false negative (FN) rate, defined as the probability that ANOVA fails to locate steps with disturbances.

Nonuniform transition probabilities can also increase the false positive (FP) rate, defined as the probability that ANOVA incorrectly detects a disturbance at a step. This is illustrated by the following example. Fig. 2 shows the subset distributions of lots grouped by two tools at Step 2 in the simulated process shown in Fig. 1. The only disturbance in the process is at Tool A in this step. ANOVA correctly detects the disturbance. Due to the nonuniform transition probabilities, lots processed by Tool A in Step 2 are likely to be processed by Tool A in Step 3. As shown in Fig. 3, this shifts the mean of the lots processed by Tool A more than Tools B or C at Step 3. Even though none of these tools contains a disturbance, ANOVA detects the difference in the means of the tool subsets and incorrectly indicates there is a disturbance at this step (i.e., generates an FP error). Thus, the practical limitations of semiconductor manufacturing cause ANOVA to generate more FN and FP errors than would be expected if the tool usage and transition probabilities were uniform [2].

We propose two new methods to locate disturbances based on stepwise regression (SWR). The first is called step-at-a-time stepwise regression (S-SWR). Like ANOVA, S-SWR is designed to locate *steps* in the process with disturbances. The second method is called step-tool stepwise regression (ST-SWR). Unlike ANOVA and S-SWR, ST-SWR is designed to locate *tools* in the process with disturbances. We designed both SWR methods to consider the entire process during the analysis, rather than analyzing each step individually like ANOVA. This gives these methods the ability to account for the nonuniform transition probabilities that confound ANOVA.

Although ANOVA is the primary method used to locate disturbances in practice, several alternatives have been proposed. For example, tool commonality analysis can identify some tools that decrease yield [3]. Unlike ANOVA and the SWR methods, this analysis only applies to categorical data and only uses data from bad lots to locate disturbances. The binomial test can be used to detect when disturbances exist, but does not locate where the disturbances are in the process [4]. Multi vari analysis and ANOVA can be combined to detect changes in the process, but this is used on individual tools and cannot be used to locate disturbances in the entire process with one analysis [5]. Linear regression can also be used to locate tools between inspection points that generate defects, but this requires many more lots than ANOVA and the SWR methods [1].

An earlier version of the S-SWR was first described in [6]. The proposed stopping rule for this algorithm causes a higher rate of false positives (FPs) than the new stopping rule proposed here and did not have a user-specified parameter for controlling the level of significance [7]. The S-SWR algorithm described in this paper overcomes both of these problems.

This paper makes several new contributions to help address the problem of locating disturbances. First, we frame the problem as a type of subset selection. Second, we describe two new SWR methods that are data efficient and take account of the nonuniform transition probabilities that confound methods that analyze steps independently such as ANOVA. Third, we describe a novel method of statistically modeling semiconductor manufacturing processes that can be used to assess the effectiveness of these methods. Finally, we perform a thorough assessment of ANOVA and the SWR methods with Monte Carlo simulations over a wide range of realistic processing conditions.

## II. Algorithm Design

### A. Subset Selection

Fundamentally, the problem of locating disturbances can be framed as a subset selection problem where the goal is to identify subsets of lots that are statistically distinct from other subsets. A different subset can be defined for each sequence of step-tool combinations in the process.

The natural analysis method for this problem is $n_s$-way ANOVA, where $n_s$ is the number of steps in the process. In

this application, however, the data sets are often too small to divide into all possible subset combinations. A process with $n_s$ process steps and $n_t$ tools at each step would require forming

$$n_p = \sum_{k=1}^{n_s} \binom{n_s}{k} n_t^k \tag{1}$$

different subsets. For a process with 200 steps and five tools at each step, this works out to be $> 10^{155}$ possible subsets. Since the data sets available for analyzing the process typically contain only 25–100 lots, a full consideration of all possible subsets with $n_s$-way ANOVA is computationally and statistically impractical.

There are a number of suboptimal subset selection methods that can be applied to smaller data sets. Details of these methods and comparisons of their performance relative to one another can be found in [8]–[10].

### B. Statistical and Regression Models

Of these subset selection methods, SWR is one of the simplest to implement and is well suited for use with additive data. In this application, SWR is based on the following statistical model of how the data was generated:

$$y_i = \epsilon_i + \sum_{j=1}^{n_d} \beta_j I_j(i) \tag{2}$$

where $y_i$ is the $i$th observation (e.g., yield of $i$th lot), $\epsilon_i$ is an independent and identically distributed (i.i.d.) random variable, $n_d$ is the number of subsets with different means, $\beta_j$ is the contribution to the mean by the $j$th subset, and $I_j(i)$ is an indicator variable. This variable is one if the $i$th observation is included in the $j$th subset and zero otherwise. In our application, each indicator variable represents a single tool at a single step. If lots were processed by this tool at this step, the indicator variable is assigned a value of one for these lots. For other lots, this indicator variable is assigned a value of zero.

This statistical model assumes that each subset is normally distributed, that $\epsilon_i$ has constant variance, and that $\epsilon_i$ is the only source of random variation. ANOVA is based on the same statistical model and assumptions.

The SWR regression model is then defined as

$$\hat{y}_i = \sum_{j=1}^{n_d} \hat{\beta}_j I_j(i) \tag{3}$$

where $\hat{y}_i$ is the model output and the model coefficients, $\hat{\beta}_j$, are estimated from the data using linear least-squares regression.

### C. Stepwise Regression

SWR can be understood as a form of regularization that permits linear regression models to be applied to applications in which the number of input variables exceeds the number of observations. There are several variations of SWR. The version we propose for this application, called *forward stepwise regression*, begins by modeling the process as a constant: the sample mean of the observations. The model is then constructed sequentially. The algorithm first searches for the single input variable that explains the most variation. If this input increases the explained variation sufficiently, it is added to the model. This step is called *forward selection*. The algorithm then searches for the best input to remove from the model. If the removal of this input does not decrease the explained variation significantly, it is removed from the model. This step is called *backward elimination.* These two steps are repeated until no more input variables can be added or removed. A complete description of the algorithm is as follows.

**Stepwise Regression**
1. Initialize the set of variables that are included in the model, $\phi := \emptyset$
2. **Forward Selection**
   For $i = 1$ to $n_v$,
   2.1 Construct a test model that includes the input variable $i$ and all input variables in the set $\phi$
   2.2 Calculate the test model's $F_t$ statistic
   2.3 If $F_t > F$ of all prior test models in this loop, store this input variable as the best so far, $i_b := i$
3. Calculate the $F$ statistic of a model that includes the $i_b$th variable and all variables in the set $\phi$
4. If $F \geq F_a$, add $i_b$ to the set $\phi$
5. **Backward Elimination**
   For $i = 1$ to $n_\phi$,
   5.1 Construct a test model with all the variables in $\phi$ except $\phi(i)$
   5.2 Calculate the test model's $F_t$ statistic
   5.3 If $F_t < F$ of all prior test models in this loop, store this input variable as the best so far, $i_b := i$
6. Calculate the $F$ statistic of a model that includes all variables in $\phi$ except $i_b$
7. If $F < F_d$, remove $i_b$ from the set $\phi$
8. If an input variable was added in Step 4 or removed in Step 7, go to Step 2.

In this algorithm $n_v$ represents the number of possible input variables that could be included in the model, $n_\phi$ represents the number of input variables currently included in the model, $F_a$ is a user-specified threshold for *adding* an input variable to the model, and $F_d$ is a user-specified threshold for *dropping* an input variable from the model.

The $F$ statistic is one of the key components of this algorithm. This statistic is used to compare two models. During forward selection the *reduced* model contains all of the input variables currently in the model and the *full* model contains these variables and the input variable under consideration. During backward elimination the *reduced* model contains all of the variables currently in the model except the variable under consideration

and the *full* model contains all of the variables currently in the model. For both cases, the $F$ statistic is then defined as

$$F = \frac{\dfrac{\mathrm{SSE}_r - \mathrm{SSE}_f}{\mathrm{df}_r - \mathrm{df}_f}}{\dfrac{\mathrm{SSE}_f}{\mathrm{df}_f}} \quad (4)$$

where $\mathrm{SSE}_r$ and $\mathrm{SSE}_f$ are the sum of squared errors of the reduced and full models, respectively. Similarly, $\mathrm{df}_r$ and $\mathrm{df}_f$ are the degrees of freedom of the reduced and full models, respectively. The degrees of freedom of a model are equal to the difference of the number of observations, $n_o$, and the number of variables in the model, $n_\phi$. In (4) the difference $\mathrm{df}_r - \mathrm{df}_f$ is called the numerator degrees of freedom and $\mathrm{df}_f$ is called the denominator degrees of freedom. SSE is the sum of squared errors, defined as

$$\mathrm{SSE} = \sum_{i=1}^{n_o} (y_i - \hat{y}_i)^2 \quad (5)$$

where $y_i$ is the $i$th observation and $\hat{y}_i$ is the $i$th model output given by (3).

### D. Step-Tool Stepwise Regression (ST-SWR)

We call the process of applying stepwise regression to this problem step-tool stepwise regression (ST-SWR). The primary information produced by ST-SWR is the list of input variables included in the model. These represent the step-tool combinations that explain the most variance. This list represents the locations of probable disturbances in the process.

### E. Step-at-a-Time Stepwise Regression (S-SWR)

Step-at-a-time stepwise regression (S-SWR) is an unconventional application of stepwise regression. Instead of considering single step-tool combinations during forward selection and backward elimination, S-SWR considers all of the tools at each step simultaneously. This requires some modifications to the canonical SWR algorithm that is designed to add or drop a single input variable during each iteration through the main loop. These modifications are described in the following section.

Like ANOVA, the primary information produced by S-SWR is a list of manufacturing steps that explain the most variance. This list can help process engineers locate which manufacturing steps are likely to contain disturbances in the manufacturing process.

### F. Stopping Rules

SWR uses a stopping rule to determine when to stop testing variables for addition to or removal from the model. While there are several stopping rules that could be used [11], the $F$ test stopping rule described earlier is the most common. This stopping rule requires two user-specified thresholds: one for adding variables to the model, $F_a$, and one for dropping variables from the model, $F_d$. During forward selection, if the largest $F$ statistic among all the test models is greater than or equal to $F_a$, then the corresponding variable is added to the model (Step 4 in the algorithm). During backward elimination, if the smallest $F$ statistic among all the test models is less than $F_d$, then the corresponding variable is dropped from the model (Step 7 in the algorithm). The process of adding and dropping variables continues until none of the remaining variables have an $F$ value larger than $F_a$ and none of the variables in the model have an $F$ value smaller than $F_d$.

While a statistical hypothesis test is used to add or drop variables, the $F$ statistics only approximately follow an $F$ distribution [12]. This prevents the calculation of meaningful $p$ values for adding or dropping variables. Despite these deficiencies, the $F$ test is the most popular stopping rule for stepwise regression and has worked well in many practical applications.

### G. New Stopping Rule for S-SWR

Normally the user-specified $F_a$ and $F_d$ thresholds, or their corresponding $p$-values, are specified at the beginning of the SWR process. This is possible with canonical SWR because only one input variable is considered during forward selection and backward elimination. Thus, there is one degree of freedom in the numerator and approximately $n_o$ degrees of freedom in the denominator. This approximation to the true degrees of freedom, $n_o - n_\phi$, is acceptable in most applications because the number of observations is usually much larger than the number of variables in the model $(n_o \gg n_\phi)$.

S-SWR considers whether to add groups of input variables during each forward selection iteration and whether to remove groups of input variables during each backward elimination iteration. Each group includes indicator variables for all of the tools at a single step under consideration. Since the numerator degrees of freedom are equal to the number of tools in the step under consideration, the distribution of the $F$ statistic varies and fixed thresholds $F_a$ and $F_d$ are inappropriate. We overcome this problem by replacing the forward selection and backward elimination steps with more traditional hypothesis tests and specify thresholds in terms of $p$ values that are based on the correct degrees of freedom.

**Step-at-a-time Stepwise Regression (S-SWR)**

1. Initialize the set of steps that are included in the model $\varphi := \emptyset$
2. **Forward Selection**
   For $i = 1$ to $n_s$,
   2.1 Construct a test model that includes the input variables representing all of the tools at step $i$ and all of the steps in the set $\varphi$
   2.2 Calculate $p_t$, the test model's $p$ value
   2.3 If $p_t \le p$ of all prior test models in this loop, store this step as the best so far, $i_b := i$
3. Calculate the $p$ value of a model that includes input variables representing all of the tools at step $i_b$ and the steps in the set $\varphi$

```
4. If p ≤ p_a, add the step i_b to the set φ
5. Backward Elimination
 For i = 1 to n_φ,
 5.1 Construct a test model with all the
 steps in φ except those in step i
 5.2 Calculate p_t, the test model's p
 value
 5.3 If p_t > p of all prior test models in
 this loop, store this step as the best so
 far, i_b := i
6. Calculate the p value of a model that
 includes all input variables for the
 steps in φ except step i_b
7. If p > p_d, remove step i_b from the set φ
8. If a variable was added in Step 4 or
 removed in Step 7, go to Step 2.
```

The use of $p$ values for thresholds instead of $F$ statistics has the additional benefit that the user can loosely interpret the thresholds as levels of significance (i.e., false positive rate) that are commonly used with ANOVA and other hypothesis tests. Typical values are $p_a = p_d = 0.05$ and 0.01.

### *H. Multicollinearity and Regularization*

To calculate the model coefficients in (3) that minimize the sum of squared errors, the input variables are typically collected into a single matrix, $\boldsymbol{A}$. One row is used for each observation and each column represents the value of the corresponding input variable. Similarly, the lot statistic of interest (e.g., yield or defectivity) can be collected into a single vector, $\boldsymbol{y}$. Each element of the vector contains the value of the lot characteristic for each observation. The least squares estimate of the model coefficients can be found by solving the normal equations

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} = \boldsymbol{y} \tag{6}$$

where $\hat{\boldsymbol{\beta}}$ is a vector containing the model coefficients in (3) and $\boldsymbol{y}$ is a vector containing the observed statistics for each lot modeled by (2). If $\boldsymbol{A}$ has full column rank, then the least squares solution for the model coefficients is given by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}. \tag{7}$$

If $\boldsymbol{A}$ does not have full column rank, the columns of $\boldsymbol{A}$ are not linearly independent and $\boldsymbol{A}$ is said to be collinear. This makes it difficult to solve the normal equations because the product $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is singular and does not have an inverse.

Although SWR normally prevents collinear variables from entering the model, it is not prevented with the S-SWR variation of SWR proposed here. This is because each set of input variables for a step has exactly one column with a 1 representing the tool that was used to process the lot at that step. Thus, the columns for a step sum to 1 in every row of $\boldsymbol{A}$. If a second step is added, the columns representing that step will also sum to 1 causing the matrix to be collinear. Thus, with S-SWR, collinear variables must be allowed to enter the model.

To eliminate the influence of multicollinearity, we calculate the model coefficients using principal components regression (PCR). This is based on a singular value decomposition of $\boldsymbol{A}$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} \tag{8}$$

where $\boldsymbol{U} \in \mathbb{R}^{n_o \times n_o}$ is a unitary matrix, $\boldsymbol{V} \in \mathbb{R}^{n_\phi \times n_\phi}$ is a unitary matrix, and $\boldsymbol{\Sigma} \in \mathbb{R}^{n_o \times n_\phi}$ is a diagonal matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are called the singular values of $\boldsymbol{A}$, and they are normally arranged in decreasing order so that $\sigma_i \geq \sigma_j$ for all $i < j$. The vector of coefficients can then be written as

$$\hat{\boldsymbol{\beta}} = \sum_{i=1}^{n_\phi} \frac{\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{y}}{\sigma_i}\boldsymbol{v}_i \tag{9}$$

where $\boldsymbol{u}_i$ is the $i$th column of $\boldsymbol{U}$, $\boldsymbol{v}_i$ is the $i$th column of $\boldsymbol{V}$, $\sigma_i$ is the $i$th diagonal element of $\boldsymbol{\Sigma}$, and $n_\phi$ is the number of input variables included in the model. If $\boldsymbol{A}$ is collinear, then one or more singular values will be zero. PCR merely truncates this sum so as to exclude minor components with singular values that are too small. If only $n_\sigma$ terms are included in (9), then the degrees of freedom of the reduced and full models is given by

$$\mathrm{df} = n_o - n_\sigma. \tag{10}$$

The $p$ values can then be calculated from the inverse of the $F$ distribution in the usual way. For the results reported here, we excluded terms in (9) with a singular value less than 3.16% of the maximum singular value, $\sigma_{\max}$.

## III. Design of Algorithm Assessment

We used Monte Carlo simulations to assess the ability of ANOVA and the new SWR methods to locate disturbances in a semiconductor manufacturing process. These were based on statistical models of four actual manufacturing processes at LSI Logic's fabrication facility in Gresham, Oregon. The results of simulations from one of these processes are included in this paper. The results of simulations of the remaining processes can be found in [7]. These models were composed of the same manufacturing steps and tools as the actual processes. The effect of each manufacturing step on a lot was modeled as additive Gaussian noise. The mean and variance of the distribution for each tool at each step were scaled so that the final distribution at the end of the process was the same as the actual yield. The transition probabilities between steps were estimated from data sets of the actual process. This ensured that the tool usage at each step was the same as the actual manufacturing process. Disturbances were modeled as having an additional, constant additive effect on lots that were processed by specific tools.

The following sections describe simulation parameters that we varied to assess the SWR methods' ability to locate disturbances over a range of different manufacturing conditions. These simulation batches have some properties in common. For every set of simulation parameters, we created 1000 separate realizations, or simulated data sets. This enabled us to accurately estimate the false positive (FP) and false negative (FN) error rate of each method. When the simulated process included one or more disturbances, the size of the disturbance is expressed

in terms of the standard deviation of the output of a disturbance-free process. The first four simulations, described in Sections III-A–D used only 25 steps of the manufacturing process. These steps were randomly selected from the entire simulated process for each of the 1000 simulations. These roughly correspond to the number of steps between inspection steps in the actual manufacturing process.

### A. Number of Lots

The purpose of these simulations was to assess the variability of the methods' FP rates as $n_o$, the number of lots in the data set, varied. There were no disturbances included in this batch of simulations, so any step or tool that was indicated as having a disturbance by any of the methods was an FP error.

Since these techniques are based on statistical hypothesis tests, ideally the FP rate will not be affected by $n_o$. We varied the number of lots processed during each of the 1000 simulations from 20 to 200.

### B. Number of Lots With One Disturbance

This purpose of these simulations was to assess the ability of the methods to detect a single disturbance. The design of this batch is identical to the previous one, except each simulation included one disturbance. The size of the disturbance was always $5\sigma$, where $\sigma$ is defined as the standard deviation of the data produced by an undisturbed process. This batch of simulations tested each method's ability to detect one disturbance in 25 steps. Since there was always one disturbance in the process, one FN error could occur during each of the 1000 simulations.

### C. Number of Disturbances

The purpose of these simulations was to determine how the methods' sensitivity was affected by multiple disturbances. In this batch of simulations we varied the number of disturbances from 1 to 10. Only one disturbance was placed at any given step. The number of lots was fixed at 25 and each disturbance had a value of $5\sigma$.

### D. Two Disturbances

The purpose of these simulations was to compare the specificity of the SWR methods with ANOVA. Each simulation in this batch included two disturbances. One was held constant at $5\sigma$ while the second varied from 0 to $5\sigma$. The number of lots was fixed at 25. As many as two FN errors could occur during any simulation.

### E. Number of Steps

The purpose of these simulations was to assess the ability of the methods to detect disturbances with measurements obtained from the end of the manufacturing process. We varied the number of steps from 25 to the maximum number of steps in each of the processes being simulated. The number of lots was fixed at 50. Each simulation included two disturbances: one with an amplitude of $1\sigma$ and another with an amplitude of $5\sigma$. Thus, up to two FN errors could occur during each simulation.

### F. Performance Metrics

During each simulation, we counted the number of times one or more FP errors occurred and the number of times one or more FN errors occurred. Since each of the simulations was statistically independent, the total number of times one or more errors occurred has a binomial distribution. This permitted us to find the maximum likelihood estimates of the probability of one or more FP and FN errors and the 2.5%–97.5% confidence interval of each probability estimate. This confidence interval is shown by the thin vertical lines in the following figures.

Since ST-SWR locates tools with disturbances, instead of steps with disturbances, as does ANOVA and S-SWR, we had to diminish its precision and merely decide whether it located one or more disturbances at a step that actually contained a disturbance. Thus, for all of the methods an FP error was counted for each step that was indicated as having one or more tool disturbances that actually did not have one. Similarly, an FN error was counted for each step that was not indicated as having one or more tool disturbances that actually did have one.

Each of the three methods included in the study had a single user-specified parameter, $\alpha$, that is approximately equivalent to the level of significance, or expected rate of false positives, in a hypothesis test. We report the results for ANOVA with a typical value of $\alpha = 5\%$. We chose $p_a = p_d = 1\%$ for S-SWR. We selected a smaller equivalent $p_a = p_d = 0.25\%$ for ST-SWR because this method has a much larger pool of possible input variables to include in the model: one for each step-tool combination. Thus smaller $p_a$ and $p_d$ are necessary to make the performance of ST-SWR comparable to the other methods.

Ideally we would like to adjust $\alpha$ for each of the three methods so that each had the same FP rate. We could then compare the techniques based on their sensitivity to disturbances alone. Unfortunately, these are not canonical hypothesis tests and it is not known how the expected FP rate is related to $\alpha$ for such a complex application. We have intentionally chosen a value of $\alpha$ for ANOVA that results in a larger FP rate than is obtained by the SWR methods. Because this also increases the sensitivity of ANOVA, this enables us to interpret the FN rate as a lower bound on the FN rate that would have been obtained if ANOVA had exactly the same FP rate as either of the SWR methods. Thus, if a SWR method has *both* a smaller FP rate and a smaller FN rate than ANOVA, we can conclude that the SWR method would perform better even if the FP rate of ANOVA was chosen to be the same as the SWR method. Thus, if SWR has better FP and FN rates, we conclude that the method is better regardless of the level of significance. Results for ANOVA with $\alpha = 1\%$ can be found in [7].

## IV. Results

### A. Number of Lots

Fig. 4 shows the results of a simulation in which the number of lots was varied from 20 to 200. These results show that all of the methods maintain a relatively constant FP error rate regardless of the number of lots in the data set. These results do

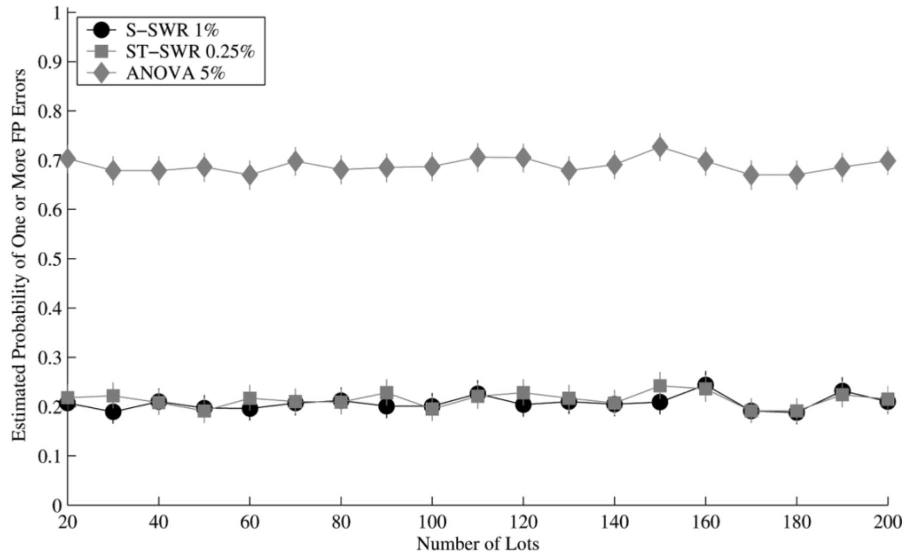
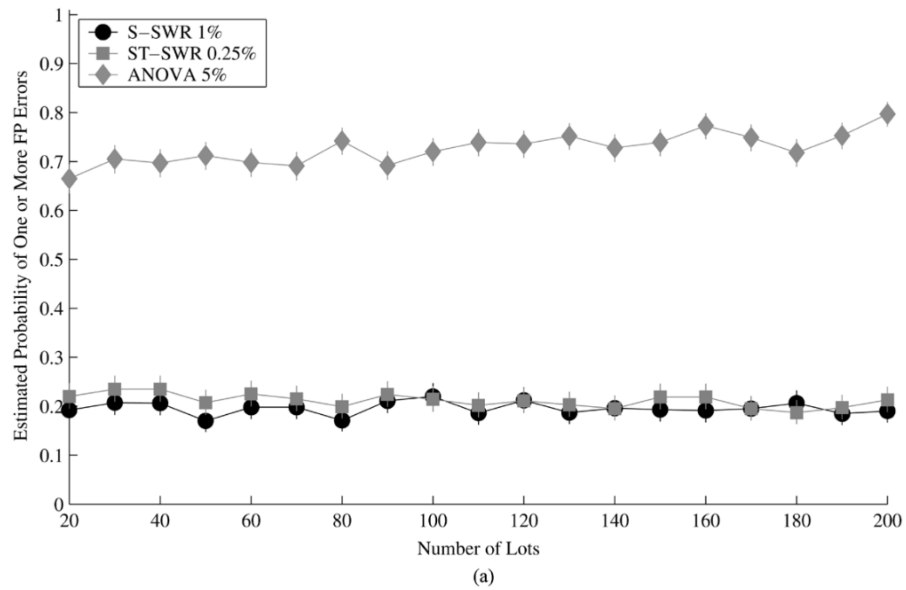
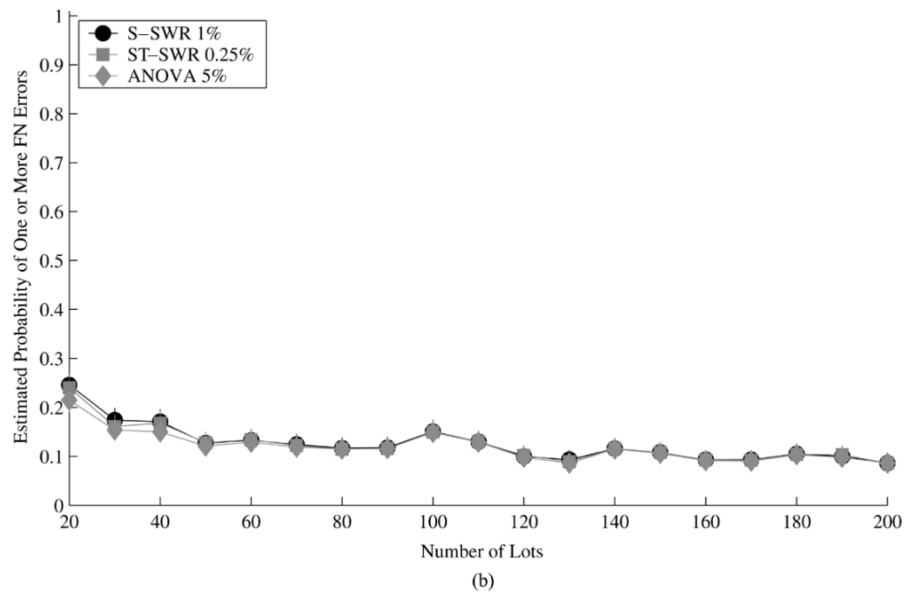



Fig. 4. Probability of one or more FP errors in a process with no disturbances. The number of steps was fixed at 25 and the number of lots varied.




(a)




(b)

Fig. 5. Estimated FP and FN error probabilities. The processes in these simulations included 1 disturbance, a varying number of lots, and 25 steps. (a) Estimated probability of one or more FP errors. (b) Estimated probability of one or more FN errors.

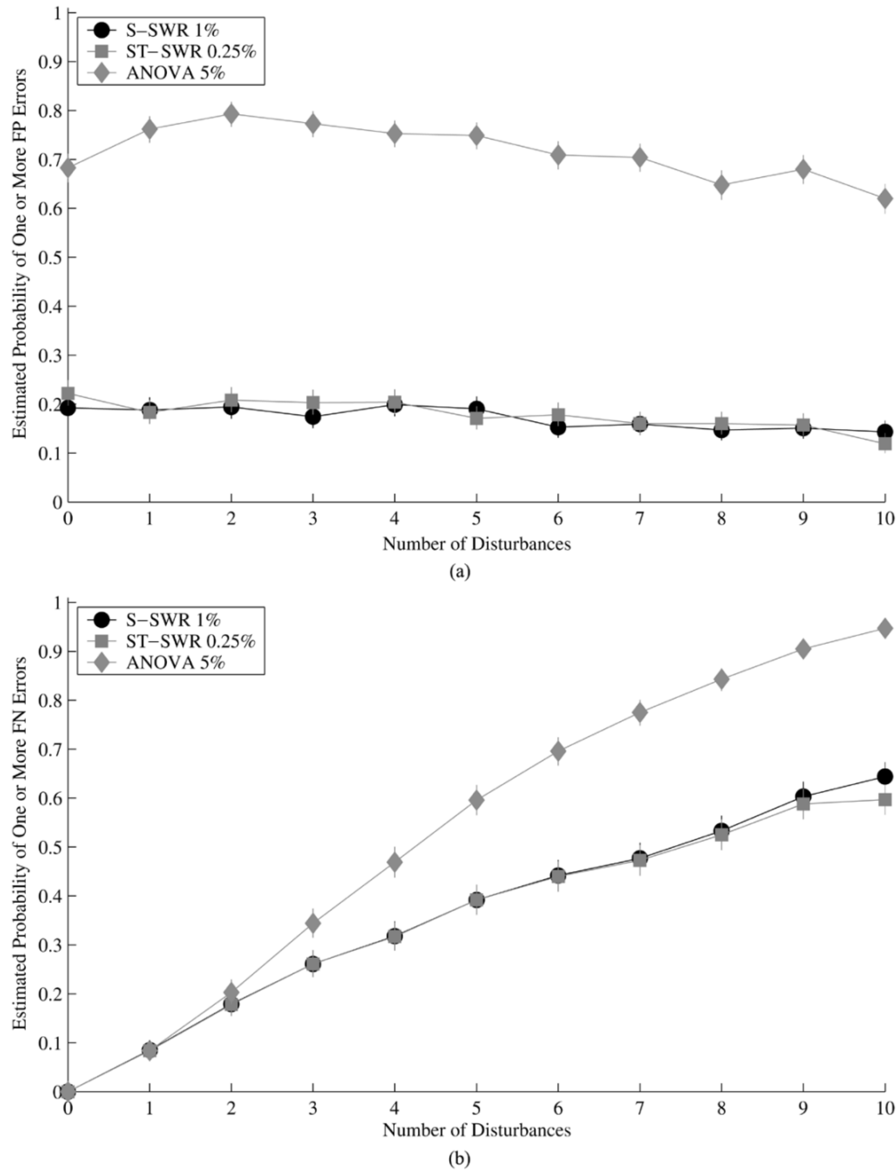



Fig. 6. Estimated FP and FN error probabilities. The processes in these simulations included a varying number of disturbances, 25 lots, and 25 steps. (a) Estimated probability of one or more FP errors. (b) Estimated probability of one or more FN errors.

not give any information about which methods perform best because they give no indication of how sensitive the methods are to detecting disturbances.

### *B. Number of Lots With One Disturbance*

Fig. 5 shows the FP and FN error rates for a process with one disturbance. Both of the SWR methods have FP error rates that are much lower than ANOVA, and all three methods have nearly the same FN error rate. Thus the SWR methods perform better than ANOVA.

### *C. Number of Disturbances*

Fig. 6 shows the FP and FN error rate over a range of disturbances. Like the results in the previous section both SWR methods have lower FP error rates than ANOVA. In these simulations, however, the SWR methods also have lower FN error rates than ANOVA. Thus, both SWR methods significantly outperform ANOVA in these simulations.

### *D. Two Disturbances*

Fig. 7 shows the FP and FN error rate over a range of amplitudes of a second disturbance. As with the previous batch of simulations, both SWR methods significantly outperform ANOVA.

### *E. Number of Steps*

Fig. 8 show the FP and FN error rates for simulated processes with a varying number of processing steps. Again, the SWR methods significantly outperform ANOVA. In this case, the FN rate is lower for all of the techniques than in the previous simulations because 50 lots were used instead of 25.

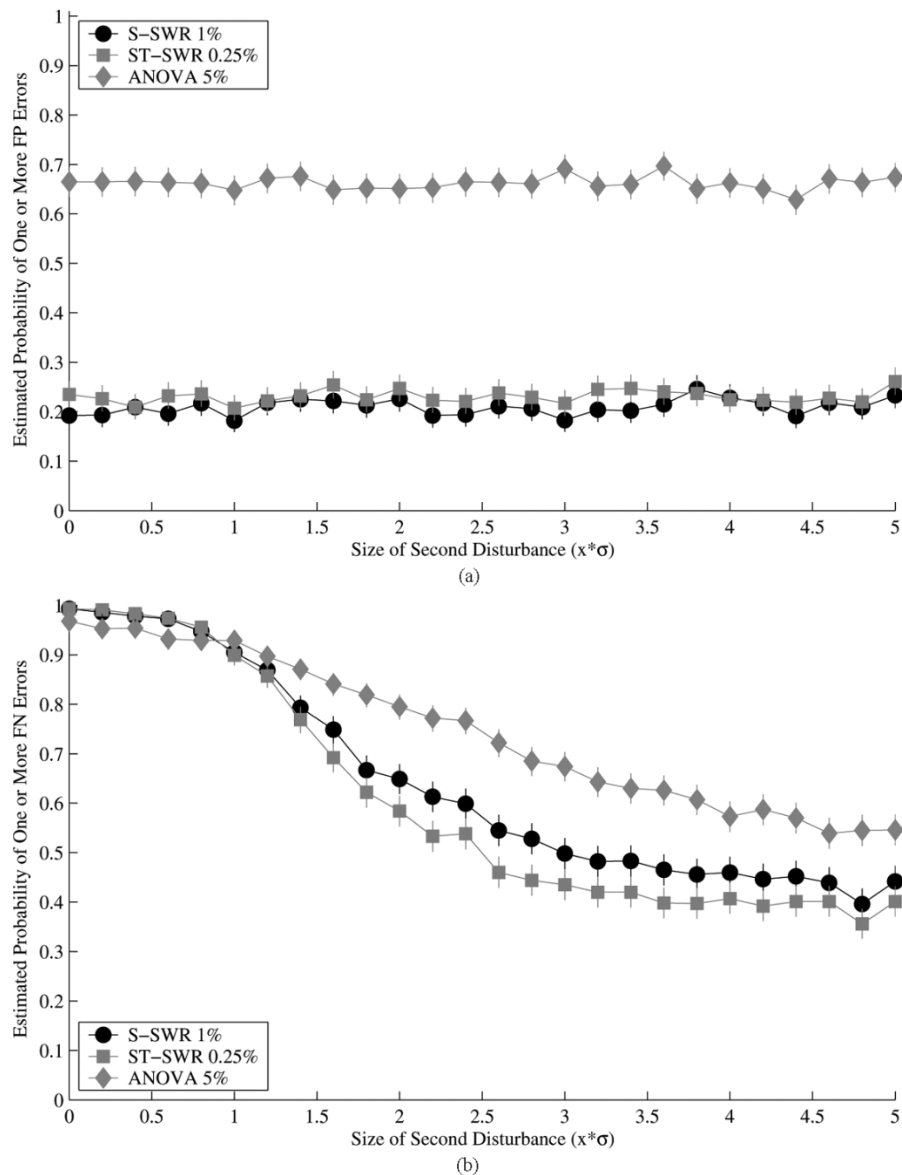



Fig. 7. Estimated FP and FN error probabilities. The processes in these simulations included two disturbances, 25 lots, and 25 steps. (a) Estimated probability of one or more FP errors. (b) Estimated probability of one or more FN errors.

## V. Discussion

Although the simulation model accounted for the nonuniform tool usage and transition probabilities between manufacturing steps, it does not account for tool drift, maintenance cycles, and outliers that occurs in practice. These effects cause the observations to be correlated over time and degrade the performance of all three methods. Nonetheless, the relative performance of these methods is likely to remain the same in real manufacturing processes.

Throughout all five batches of simulations both SWR methods consistently outperformed ANOVA. The disadvantages of ANOVA, the best current practice, are exemplified by the results in Fig. 8. For a process with 140 steps, ANOVA with a 5% level of significance has nearly a 100% probability of generating one or more FP errors and a 75% probability of generating one or more FN errors. This means that for any process with more than 140 steps, which is essentially all modern semiconductor manufacturing processes, ANOVA will generate at least one false alarm every time it is used to analyze the process while failing to detect all of the disturbances 75% of the time.

The SWR methods perform much better. They achieve an FP error probability of approximately 80% and an FN error probability of approximately 55%–65%. Both of these probabilities are lower than ANOVA. Thus, the SWR methods generate fewer false alarms and detects actual disturbances more often.

### A. Advantages and Limitations of SWR

The primary advantage of the new SWR methods as compared to ANOVA is that they are able to consider the entire process as a holistic problem, rather than analyzing each step independently of the other steps. When an input variable is added to the regression model with SWR, the variance associated with it is essentially removed from the variance of the output data that the remaining input variables try to explain. This enables the distributions of the remaining variables to be estimated more accurately and minimizes the confounding caused by nonuniform mixing.

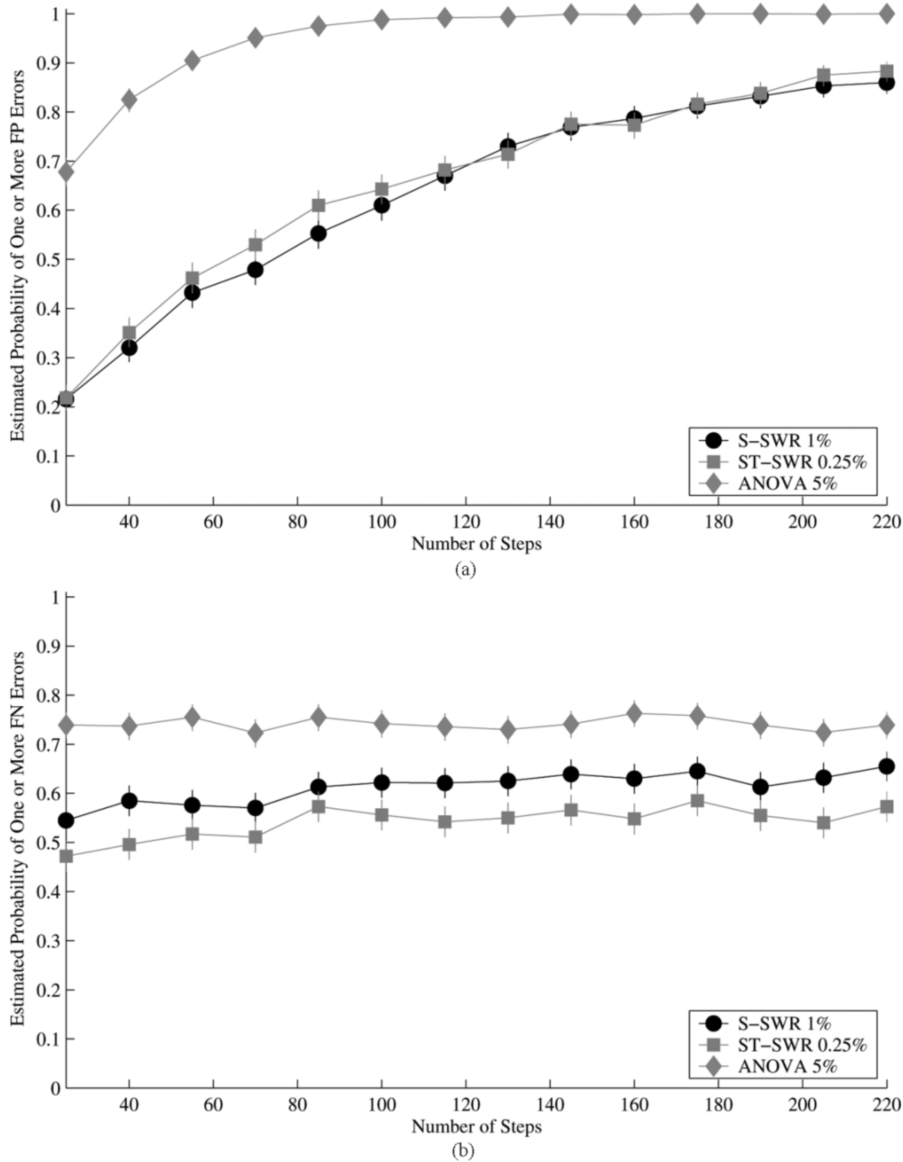



Fig. 8. Estimated FP and FN error probabilities. The processes in these simulations included two disturbances, 50 lots, and a varying number of steps. (a) Estimated probability of one or more FP errors. (b) Estimated probability of one or more FN errors.

To demonstrate this advantage consider using S-SWR to analyze the process shown in Fig. 1. A simulation of this process results in the distributions shown in Figs. 2 and 3 for Steps 2 and 3. S-SWR finds that Step 2 is the most significant and places it in a model. With Step 2 in the model, the variance associated with it is essentially removed from Step 3. The resulting residual distributions for the tools in Step 3 are shown in Fig. 9. The distributions appear more normal, and they have nearly equal averages. S-SWR *correctly* does not detect a disturbance at this step. Thus while ANOVA generated a FP error at this step due to nonuniform mixing, S-SWR accounts for the nonuniform mixing and correctly decides there is no disturbance at this step.

Another advantage of SWR over ANOVA is its ability to remove a step from the model that had previously been added. It is possible that, on the first pass through the variables, the most significant variable only appeared significant due to the problems caused by nonuniform mixing. Once the steps that are truly responsible for the disturbances are identified, the first step that was added to the model would likely no longer appear significant and could be removed. Thus even if SWR initially

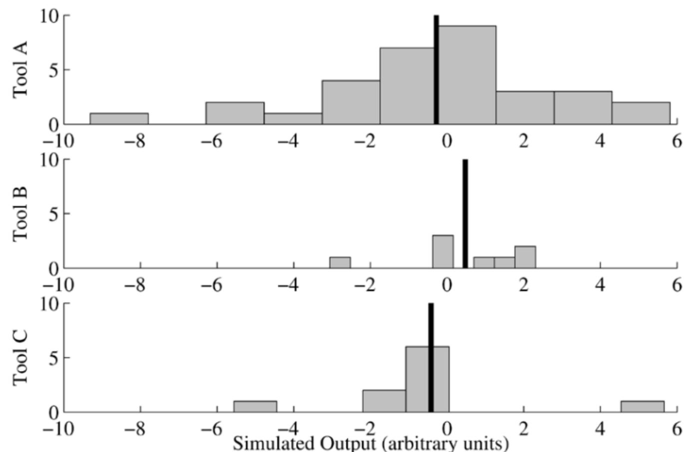



Fig. 9. Histograms of the residuals after the disturbance in the process in Fig. 1 at Step 2 is accounted for by SWR. In this case, there is no significant difference between the averages of the distributions and SWR correctly decides there is no disturbance at this step.

added an incorrect step, it is possible that it would remove it later. ANOVA does not have this capability.

The combination of principal components regression (PCR) and SWR may be novel. Canonical SWR prevents collinear steps from entering the model, so collinearity is normally not a problem. In this application, however, collinear variables must be allowed into the model. PCR makes this possible.

## VI. Conclusion

We introduced two new methods for locating disturbances in semiconductor manufacturing processes. S-SWR is used to build a linear model of the process by considering whether each step can explain a significant portion of the variance. ST-SWR is used to build a linear model of the process by considering whether each step-tool combination can explain a significant portion of the variance.

Results of over 90 000 Monte Carlo simulations suggest that these new SWR methods locate disturbances with fewer false positives and false negatives than ANOVA, the best current practice. In practical terms, this means process engineers will spend less time responding to false alarms and will be able to locate real disturbances more often.

## References


[1] K. Fridgeirsdottir, R. Akella, M. Li, P. McNally, and S. Mittal, "Statistical methodology for yield enhancement via baseline reduction," in *Proc. 1998 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, pp. 77–81.

[2] D. Turner, D. Abercrombie, J. McNames, R. Daasch, and R. Madge, "Isolating and removing sources of variation in test data," in *Proc. 2002 Int. Test Conf.*, pp. 464–471.

[3] G. Kong, "Tool commonality analysis for yield enhancement," in *Proc. 2002 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, pp. 202–205.

[4] U. Kaempf, "The binomial test: A simple tool to identify process problems," *IEEE Tran. Semicond. Manuf.*, vol. 8, no. 2, pp. 160–166, May 1995.

[5] L. K. Garling and G. P. Woods, "Determining equipment performance using analysis of variance," in *Proc. 1990 Int. Semiconductor Manufacturing Science Symp.*, pp. 85–89.

[6] A. T. McCray, J. McNames, and D. Abercrombie, "Stepwise regression for identifying sources of variation in a semiconductor manufacturing process," in *Proc. 2004 IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, pp. 448–452.

[7] A. T. McCray, "Stepwise regression for semiconductor manufacturing fault isolation and yield management," M.S. thesis, Portland State Univ., Portland, OR, 2004.

[8] T. Almoy, "A simulation study on comparison of prediction methods when only a few components are relevant," *Comput. Stat. Data Anal.*, vol. 21, no. 1, pp. 87–107, Jan. 1996.

[9] R. W. Hoerl, J. H. Schuenemeyer, and A. E. Hoerl, "A simulation of biased estimation and subset selection regression techniques," *Technometrics*, vol. 28, no. 4, pp. 369–380, Nov. 1986.

[10] P. A. Murtaugh, "Methods of variable selection in regression modeling," *Commun. Stat.: Simulat. Comput.*, vol. 27, no. 3, pp. 711–734, 1998.

[11] R. B. Bendel and A. A. Afifi, "Comparison of stopping rules in forward stepwise regression," *J. Amer. Stat. Assoc.*, vol. 72, no. 357, pp. 46–53, Mar. 1977.

[12] P. T. Pope and J. T. Webster, "The use of an $F$-statistic in stepwise regression procedures," *Technometrics*, vol. 14, no. 2, pp. 327–340, May 1972.

**Anthony T. McCray** (S'02) received the B.S. degree in computer engineering and the M.S. degree in electrical and computer engineering from Portland State University, Portland, OR, in 2003 and 2004, respectively.

He is a Quality Engineer with Sun Microsystems, Hillsboro, OR. His primary duties include collecting and analyzing data from Sun's server manufacturing operation and reliability testing labs, with a focus on measuring and improving product quality. He has six years of experience in server manufacturing operations, including several years driving product quality enhancements through the application of accelerated life testing.

**James McNames** (M'99–SM'03) received the B.S. degree in electrical engineering from California Polytechnic State University, San Luis Obispo, CA, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1995 and 1999, respectively.

He has been with the Electrical and Computer Engineering Department at Portland State University, Portland, OR, since 1999, where he is currently an Associate Professor. He has published over 100 journal and conference papers. His primary research interest is statistical signal processing with applications to statistical process control of semiconductor manufacturing, spatial wafer modeling, fault isolation, yield enhancement, and biomedical engineering. He is a member of the Integrated Circuits Design and Test Laboratory and director of the Biomedical Signal Processing Laboratory.

**David Abercrombie** received the B.S.E.E. degree from Clemson University, Clemson, SC, in 1987 and the M.S.E.E. degree from North Carolina State University, Raleigh, in 1988.

He is the Design for Manufacturing Program Manager at Mentor Graphics, Wilsonville, OR. He drives the roadmap for defining and developing EDA tools to solve the growing issues in design to process interactions that are creating ever increasing yield problems in advanced semiconductor manufacturing. He has 15 years of experience driving yield enhancement programs in semiconductor manufacturing at LSI Logic, Motorola, Harris, and General Electric. He has also led software development teams in delivering yield enhancement and data mining solutions to semiconductor manufacturing. He is extensively published in papers and patents on semiconductor processing and yield enhancement.