Multivariate Statistical Methods for Modeling and Analysis of Wafer Probe Test Data

Katina R. Skinner, Douglas C. Montgomery, George C. Runger, John W. Fowler, Daniel R. McCarville, Teri Reed Rhoads, and James D. Stanley

***Abstract*—Probe testing following wafer fabrication can produce extremely large amounts of data, which is often used to inspect a final product to determine if the product meets specifications. This data can be further utilized in studying the effects of the wafer fabrication process on the quality or yield of the wafers. Relationships among the parameters may provide valuable process information that can improve future production. This paper compares many methods of using the probe test data to determine the cause of low yield wafers. The methods discussed include two classes of traditional multivariate statistical methods, clustering and principal component methods and regression-based methods. These traditional methods are compared to a classification and regression tree (CART) method. The results for each method are presented. CART adequately fits the data and provides a "recipe" for avoiding low yield wafers and because CART is distribution-free there are no assumptions about the distributional properties of the data. CART is strongly recommended for analyzing wafer probe data.**

***Index Terms*—CART, multivariate statistical methods, tree regression, yield analysis.**

I. Introduction

SEMICONDUCTOR devices are typically described in terms of many electrical parameters. The number of parameters may exceed two hundred for a microprocessor device—many of which are correlated because process variation at one of the many manufacturing steps can influence several electrical parameters. Similarly, several steps may influence each parameter.

High-volume wafer fabrication facilities typically produce thousands of wafers per week, and each wafer may have from fifty to several thousand chips. Upon exiting wafer fabrication, most companies perform electrical tests, often called probe operations. Regardless of the technology or wafer size, the probe operation creates and archives extremely large data sets, whose primary function is to determine product acceptance or rejection. This is not a thorough or efficient use of either the test data or the resources expended in measuring and storing the data; a more effective use of these data sets can be made.

Since the wafer probe test data set is large, consisting of many measurements on potentially many parameters, multivariate statistical methods are appropriate for modeling and analyzing the relationships among the parameters and yield. A survey of literature reveals that certain types of statistical analysis of semiconductor data are quite common. Linear and nonlinear regression methods [2], [16] and neural networks [8] are common in analyzing existing data, and designed experiments and response surface methodologies are often used when analysts can obtain more data [4]. Various methods for partitioning process variation [1] and for process monitoring are also used [4]. These references are examples of such analyses.

This paper explores the use of two general classes of traditional multivariate techniques for this problem not necessarily discussed in the references above. The traditional methods are clustering and principal components and regression-type methods. A nontraditional method called classification and regression trees (CART) is also explored and results for these methods are compared using data from an accelerometer device.

The accelerometer is a surface micro-machined device created by stacking sacrificial and permanent layers on the surface of silicon wafers followed by removal of the sacrificial layers. Accelerometers are principally used as a sensor for automobile air-bag systems.

The data used in this study is described in Section II. Sections III and IV provide brief descriptions of the methods used. We comment on the suitability of these methods for modeling electrical parameter and yield data. References for further study are presented throughout.

II. Description of Data

The original probe test data file (after data cleaning to remove incomplete or erroneous records) had 1122 records with six identifier variables such as lot and wafer number, 23 sets of descriptive variables representing the electrical parameters, and average wafer yield, which is the response variable. The raw electrical parameter data include capacitance and resistance at certain temperatures. These 23 variables were then summarized using six statistics (average, standard deviation, median, first quartile, third quartile, and interquartile range) by wafer. As a result, the file contained no raw data per chip, but rather summaries of the electrical parameter variables. The result was a total of 138 ($=23 \times 6$) descriptive variables describing each

Manuscript received January 5, 2000; revised December 15, 2001. The work of J. W. Fowler was supported in part by the National Science Foundation under Grant DMI-9713750, Grant DMI-0085041, and Grant SRC-97-FJ-492. The work of G. C. Runger and K. R. Skinner was supported in part by the National Science Foundation under Grant DMI-9713518, Grant DMI-0085041, and Grant DMI-9900113. The work of D. C. Montgomery was supported in part by grants from the National Science Foundation.

K. R. Skinner, D. C. Montgomery, G. C. Runger, and J. W. Fowler are with Arizona State University Tempe, AZ 85287-5906 USA.

D. R. McCarville is with MindSpeed Technologies, Inc., Newport Beach,CA 92660 USA.

T. R. Rhoads is with the University of Oklahoma, Norman, OK 73019-1021 USA.

J. D. Stanley is with Motorola SPS, Tempe, AZ 85284 USA.

Digital Object Identifier 10.1109/TSM.2002.804901

TABLE I
VARIABLE DESCRIPTOR ABBREVIATIONS

Set – Descriptive Statistic for Variables	*Set Abbreviation*
Average	x
Standard Deviation	s
Median	m
First Quartile	qf
Third Quartile	qt
Interquartile Range	iq

record or wafer. The analyes described in this paper were performed at the wafer level.

We will refer to having six "sets" of data, each set corresponding to one of the six descriptive statistics mentioned above. The sets and their abbreviations are given in Table I.

The variable averages set (starting with "x") is the by wafer average of each of the 23 variables. The second set (starting with "s") is by wafer standard deviation of each of the 23 variables. The third through sixth sets are named similarly. The variable names are omitted for confidentiality reasons and are substituted with variable numbers; for example, $x2$ is the average of variable two and $s2$ is the standard deviation of variable two.

Although some of the variables in the "average" set were normally distributed, most of the variables that were expected to impact yield values were not. Figs. 1 and 2 show the descriptive statistics for the variables x_1 (average yield) and an important variable s_{11}. Obviously, the assumption of univariate normality is violated in both cases. In addition, a test proposed by [5], with SAS code by [7], also confirms that the data is far from multivariate normal. Finally, the descriptive variables were also not all independent; some were functions of other variables, and thus strong correlations were observed between certain pairs of variables. Preliminary study of these predictor variables indicated that the typical problematic issues associated with wafer probe data also exist with this data set: The data are not normally distributed and are not independent. The remainder of this paper describes the search for analysis techniques that will adequately deal with these common data issues while providing clues for improving average yield.

III. TRADITIONAL MULTIVARIATE ANALYSIS METHODS

Two main categories of traditional analysis methods were employed for this data set. The first category includes clustering and principal components. The second category includes a variety of regression methods.

A. Clustering and Principal Component Methods

These methods were used in an effort to explain the differences in yields by reducing the dimensionality and finding structure in the data. For good expositions on these techniques, see [6] or [11].

1) Cluster Analysis: Cluster analysis is usually performed to classify (or cluster) observations into groups having similar characteristics. Clustering algorithms (unlike classification algorithms) assume that the groups or clusters are not known before the analysis. The method of clustering used in this analysis is agglomerative clustering, which begins with no data points assigned to clusters and ends with all the points in one cluster. Several linkage methods were used including Ward's, single, and average linkage. Unfortunately, cluster interpretation provided virtually no useful inference regarding a relationship between the data points or the variables in the clusters.

2) Principal Components: Because of the extremely large size of the data set, principal components analysis (PCA) was used in an effort to reduce the dimensionality of the data. PCA identifies a subspace near most of the data. This subspace is ideally of much lower dimension than the observed data space. Consequently, the projection of the data onto the subspace is a lower dimensional summary or approximation of the data.

In PCA, the eigenvectors corresponding to the major eigenvalues of the sample covariance matrix are the subspace basis vectors. These eigenvectors are the coefficients for the linear combination of the observed variables that define the principal component variable or "score." For each data point, the score locates the data point along the axis of the eigenvector. Hopefully, a small set of principal components will account for a large proportion of the total variance in the original variables. The principal components can often be interpreted physically based on each components' coefficient.

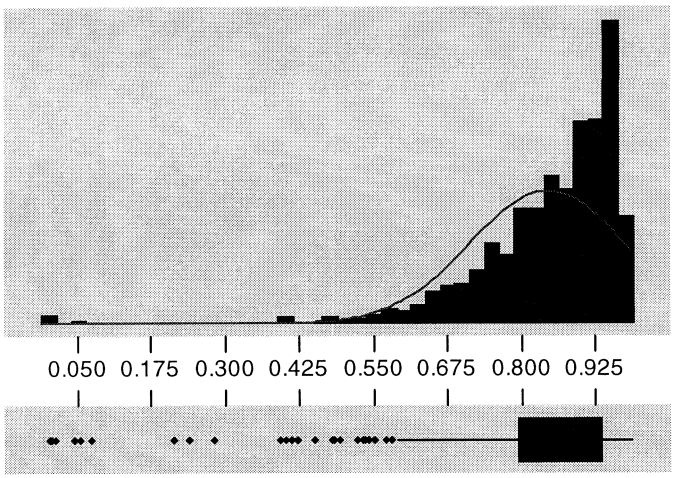
Two principal component analyses were performed in Minitab [9] using standardized variables. First, each set of variables (averages and standard deviations) were analyzed separately. Second, all six sets were combined together and principal components for these 138 variables were computed.

Both avenues produced similar results: the first five principal components explained about 60%–80% of the variation in the 138 variables. Truncated results of the PCA with all six sets are shown in Table II. Fig. 3 is a scree plot of the principal components for the combined data set. Notice that about five principal components are required to account for a reasonable proportion of the variability in the data.

To illustrate the interpretability of the principal components, we found that, for the average set of variables, the first two principal components might be interpreted as capacitance. The third principal component represents both gain and capacitance. The components for the combined data set were more ambiguous and did not indicate any obvious future course of action that would lead to yield improvement. Many of the components had the same variables and some variables seemed to be unrelated to the other variables in the component. Plots of the first five principal components showed one cluster with a number of outliers. Fig. 4 is a score plot of principal component one versus principal component two and is representative of the other plots. The black dots represent low-yield wafers. We can see that although some of the low-yield wafers are outliers on the score plot, many are indistinguishable from the high-yield wafers which are shown in gray.

B. Regression-Based Analysis Methods

After relatively little success with clustering and principal components, a series of modeling efforts using linear regression analysis was made. Description of the general procedure and the variations discussed below are in [10].

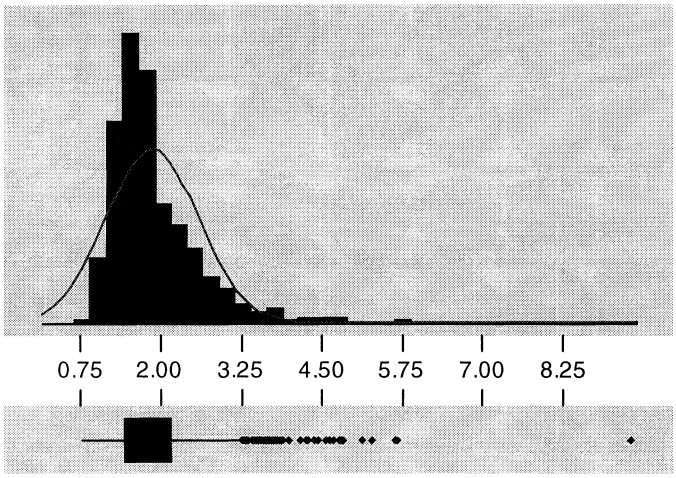


Variable: x1

Anderson-Darling Normality Test

A-Squared:	47.199
P-Value:	0.000
Mean	0.844
StDev	0.130
Variance	0.017
Skewness	-2.652
Kurtosis	11.651
N	1122
Minimum	0.000
Median	0.880
Maximum	0.985

Fig. 1. Descriptive statistics for the variable x_1 (yield) are shown. It is easy to see that the yield response variable is not univariate normal.



Variable: s11

Anderson-Darling Normality Test

A-Squared:	43.798
P-Value:	0.000
Mean	1.862
StDev	0.716
Variance	0.513
Skewness	2.532
Kurtosis	13.664
N	1122
Minimum	0.740
Median	1.680
Maximum	9.310

Fig. 2. Descriptive statistics for the variable s_{11}. Many other variables have similar attributes. It is easy to see that this variable is not univariate normal.

TABLE II
EIGENANALYSIS OF THE CORRELATION MATRIX: THE FIRST FIVE PRINCIPAL COMPONENTS

Component	*Eigenvalue*	*Proportion*	*Cumulative*
1	32.516	0.236	0.236
2	18.310	0.133	0.368
3	14.052	0.102	0.470
4	10.478	0.076	0.546
5	9.296	0.067	0.613

1) Ordinary Least Squares Regression: Ordinary least squares linear regression was used in an effort to isolate variables that might have a direct link to yield. Stepwise regression was first applied to determine potentially important variables. We obtained a model with seven important variables and an adjusted R^2 value of 0.66. The resulting prediction equation is: $yield = 500 - 0.44s_3 - 4.52s_{11} - 37.8s_{13} - 0.326iq_{15} + 1.59qt_{11} - 0.715qt_{21} - 8.07qf_{12}$, where yield is between 0 and 100. Analysis of the model residuals revealed that most of the very large residuals were associated with low yield wafers. Since we are, in effect, trying to explain low-yield situations, our model is not sufficient.

An alternative to regression on the original variables is regression on the principal components. This approach can be fruitful when the original variables are highly correlated (see [10]), as they are here. Unfortunately, linear regression on the principal components resulted in an adjusted R^2 value of only about 0.28.

Models involving transformations on the predictor variables and/or on yield, higher order polynomials terms, and spline functions are often useful (see [3] for an example of regression splines applied to semiconductor data analysis). For our example, two-way interactions of the significant variables from the original regression were created and used as candidate predictors along with the original variables in Best Subsets regression in Minitab. Important interactions were found as a result of this analysis, but the final regression equation did not account for more variability in yield than the original model which had an adjusted R^2 of 0.66.

The interaction model has an adjusted R^2 of 0.65 and uses three of the interactions. The final equation is $yield = 179 - 0.671x_{13}s_{11} - 0.425s_3 - 6.72x_{13}s_{13} - 0.33iq_{15} - 5.99iq_{21} -$

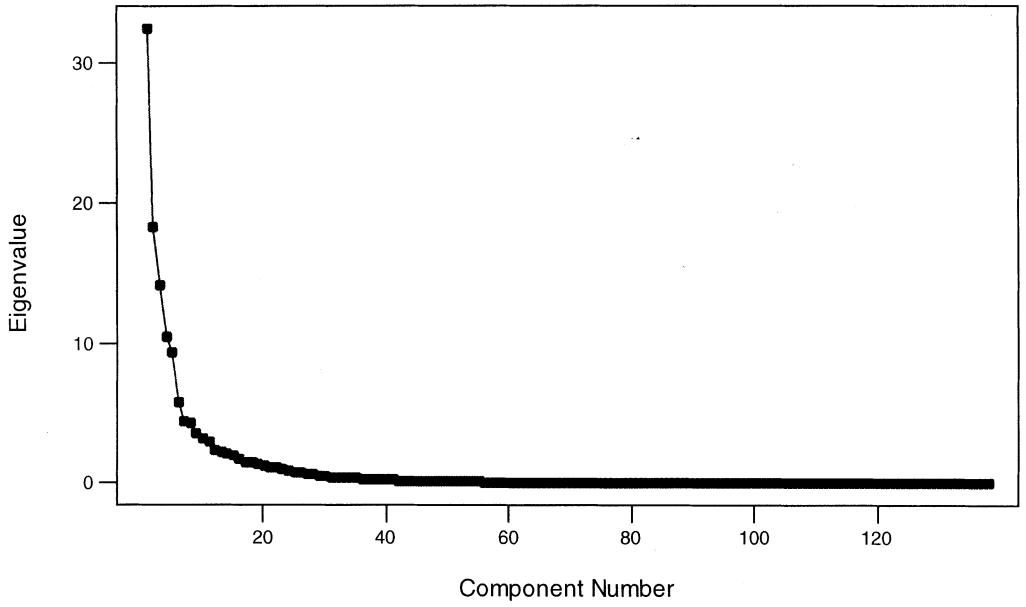


Fig. 3. Plot of eigenvalue versus principal component number is shown. Large eigenvalues (usually greater than one) explain a large amount of variation in the data. From this plot, we can see at least five principal components are required to reduce the eigenvalues to only less than ten. It would take many principal components to explain an adequate amount of the variation in the data.

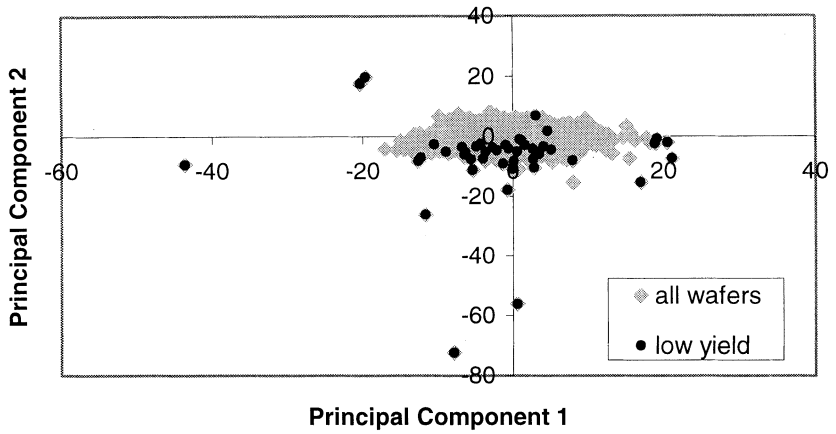


Fig. 4. Plot of first two principal components. We can see that although some of the low-yield wafers are visible outliers, more than 30 of the approximately 50 low-yield wafers are in the center of the cluster.

$7.64qf_{12} - 0.927x_{13}iq_{21}$. Plots of residuals versus the fitted yields plots reveal outliers when trying to predict yields lower than about 60%, which indicates that we still do not have a model that will adequately predict all levels of yield.

2) Generalized Linear Models: The generalized linear model (GLM) is a regression-type model that is very effective in dealing with nonnormal data. A generalized linear model consists of three components: a random component that is the response variable distribution (or error structure), a systematic component that is the linear predictor portion of the model (analogous to the expectation function in a linear regression model), and the link function between the random and the systematic components that defines the relationship between the mean of the ith observation and its linear predictor. The response distributions in a generalized linear model are all members of the exponential family, which includes the normal, gamma, exponential, Poisson, and binomial distributions. The GLM parameters are estimated by the method of maximum likelihood, so if the error distribution is normal and an identity function is specified for the link, then the GLM reduces to a classical linear regression model. See [10], [12] and [13] for more information on GLMs.

We used SAS$^{(R)}$ Proc GENMOD [14] for fitting the generalized linear model to the yield data with a binomial response distribution and the logistic link function. We performed two analyses on the combined data set: first we used the important principal components as the predictor variables, then we used the original predictor variables that were identified as important based on the stepwise regression analysis.

Wald inference statistics for the principal components model indicate that the first four components were important effects. The resulting model is

$$yield = \frac{1}{1+\exp(1.7863-0.02p_1-0.09p_2+0.08p_3-0.09p_4)}$$

where $p_1 - p_4$ are the first four principal components and yield is again between 0 and 100. An R^2-like statistic can be computed and is 0.48. Wald inference for the model using the raw variables also revealed significant effects. The resulting model is shown in the equation at the bottom of the page. It is interesting to note that the signs of variables indicate courses of action to increase yield that are opposite the actions suggested by the linear regression model. Furthermore, the signs on the model coefficients indicate that an increase in the standard deviation of several variables (3, 11, and 13) increases yield. It is possible that the attempt via the GLM (which is essentially a nonlinear model) to fit extreme groups of high- and low-yield wafers with a single model is the cause. Finally, the R^2-like statistic for this model is 0.67, which coupled with the R^2-like statistic from the principal component model, indicates that our models still do not provide a completely adequate explanation for the low yields.

C. Conclusions to Traditional Methods

The traditional multivariate methods proved unsuccessful because they either did not adequately identify conditions that would lead to high yield wafers, (as with ordinary least squares regression and generalized linear models) or because the results were difficult to interpret (as with principal components). In our experience this is a relatively common occurrence in applying these techniques to probe data; that is, it is likely not an artifact of this specific data set.

IV. Classification and Regression Trees (CART)

Another approach to model-building is based on a recursive partitioning algorithm. Predictor variables are used to partition the data points into regions with similar responses. This partitioning allows one to approximate more general response surfaces than standard regression methods. Typically a binary partitioning is used. At each step, a predictor variable is selected and a threshold is determined such that data points with the predictor below the threshold are placed in one subgroup, while the other data points form a second subgroup. The method continues by partitioning each of these subgroups by the same procedure. The method is a tree-based technique, but where most tree-based techniques are most often used with large, discrete data sets, the classification and regression tree (CART) algorithm is designed to work with continuous data. Regressors are partitioned using deviation from the average response (yield) in each subgroup as the criterion. By this approach, the user is able to build a set of paths that lead to a desired average response. For a general discussion of CART, see [2]. For an application to semiconductor manufacturing, see the dissertation of Sharma [16].

A regression tree has branches, which are paths to a leaf or node. Two important criteria are deviance and node purity. Deviance is similar to prediction error in a regression model. For this type of data, deviance $= \sum_{i=1}^{nodes} \sum_{j=1}^{n} (y_{ij} - \overline{y}_{i\cdot})^2$, where $nodes$ is the number of nodes, n is the number of points in the node and $\overline{y}_{i\cdot}$ is the average of the points in a specific node. In this case, it can also be likened to a residual sum of squares in a regression model.

Node purity measures the similarity of responses at a node (in a subgroup), and the highest node purity is with identical responses at a node. For the sake of brevity, we will not specifically mention node purity values in this paper. The goal with CART is to minimize model deviance and maximize node purity without over-fitting the model. This is accomplished by using various pruning techniques.

In addition to deviance and node purity, an R^2-like statistic can be obtained from deviance values to evaluate model fitness. R^2 alone is not a good criterion, however. The R^2 value summarizes the data across all regions, but CART users are often interested in particular regions of the data (areas of very high and very low yields). Deviance and node purity in the regions of interest are better indicators of the model fitness in such cases.

Using S-PLUS's [15] tree regression algorithm, the data were analyzed with CART. Separate trees were created using each set of predictors. We also built trees with predictors from multiple sets. The best tree included both the average and standard deviation predictors and is shown in Fig. 5. There are branches to both low and high yields.

The leftmost branch of this tree shows the best "recipe" for high yields, with average yield over 360 wafers of 93.73%. Note that the branches are conditional, which means that x_{21} less than 468.93 will do no good unless s_{11} is less than 2.23.

Deviance measures indicate that the tree explains about 62% of the deviance in the data, which is very similar to the best regression method R^2 values. This does not seem to indicate that the CART model provides benefits over the regression methods. However, as a referee has pointed out, the CART model is capable of essentially equivalent model accuracy, but may be easier to interpret and provide intuitive guidance for yield improvement. It is this ease of interpretation and intuitive guidance to improvement that make the CART model so appealing. Not only have we obtained a "recipe" for high yields, but we have identified priority variables which should be closely monitored in order to avoid very low yields.

Recall that, with the least squares regression model, the low yields appeared as outliers, indicating poor fit. The fit of the CART model in the low-yield regions has higher prediction error (as given by the "root mean square error" $= 24.5$) than the regions of high yield regions (rmse $= 5.05$). This is comparable to the least squares models where a model fit using only low-yield data has rmse $= 9.71$ and a model fit using only high-yield data as rmse $= 4.74$. However, unlike the regression model the CART model specifically identifies the low-yield regions. Further, we are not concerned with prediction in this area, but rather with identifying the variables in the area.

$$yield = \frac{1}{1 + \exp(30.32 - 0.03s_3 - 0.26s_{11} - 2.87s_{13} - 0.61qf_{12} + 0.10qt_{11} - 0.02iq_{15} - 0.05qt_{21})}$$

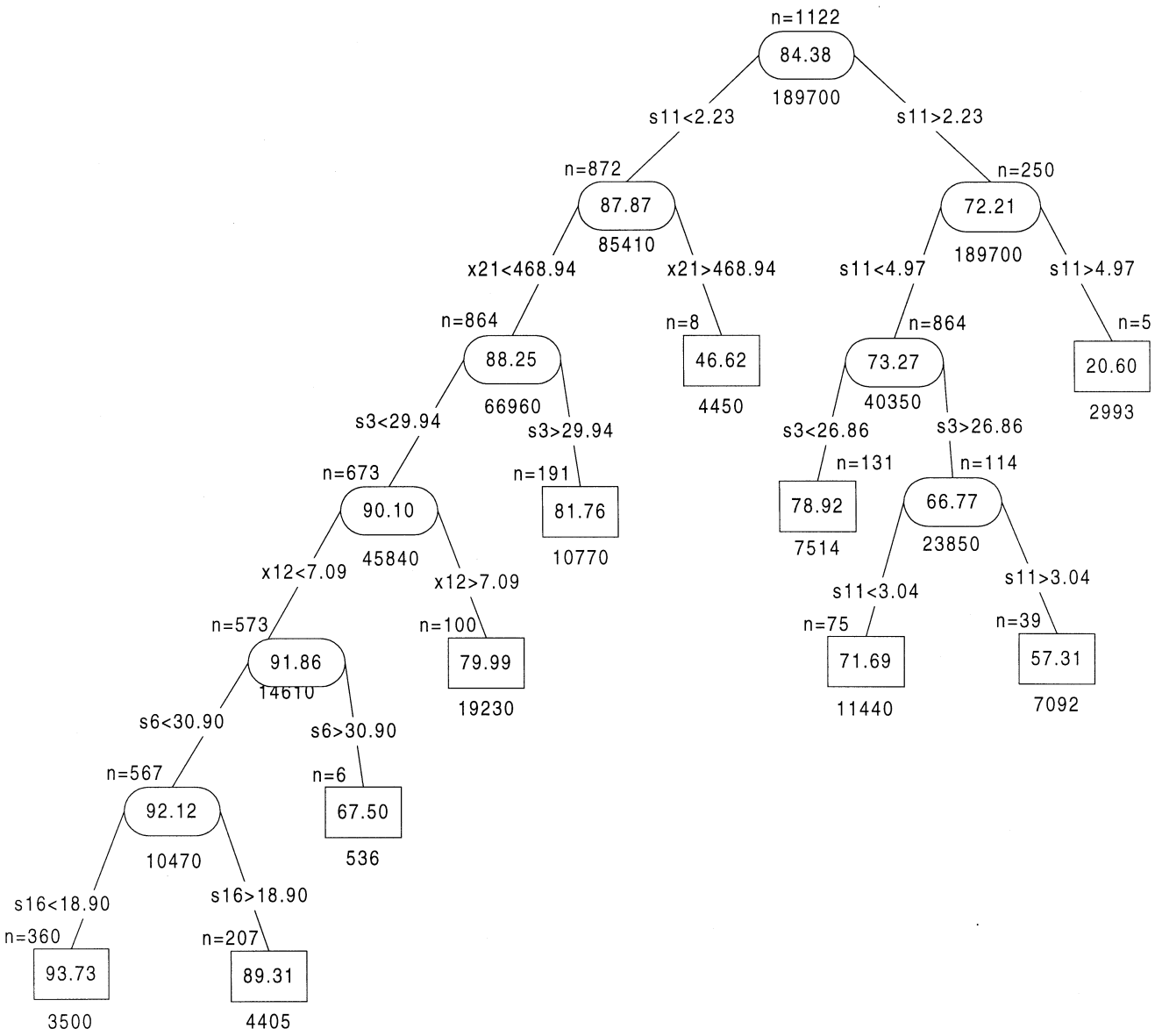


Fig. 5. Original tree with ten terminal nodes. Ovals denote nonterminal nodes. Rectangles denote terminal nodes. Deviance values are beneath the corresponding node. Sample sizes at each node are denoted by "n."

When using CART, it is best to prune the tree in order to avoid over-fitting. One way to prune the tree is by cross-validation. Cross-validation involves partitioning the data set into, say, ten subsets, forming the tree with nine, and using the results to predict the other subset. This procedure is then repeated holding out one of each of the other subsets, one at a time. A plot is produced that indicates the deviance for various sizes (number of nodes) of the tree. An example of this plot is shown in Fig. 6. More than one cross-validation analysis with various random seeds used for partitioning is usually recommended [17]. The cross-validation of the above tree indicates that about eight nodes would efficiently provide an acceptably low deviance. The resulting tree with eight nodes is shown in Fig. 7.

Fig. 7 shows that reducing the standard deviation of variable 11 (a nonlinear function of both capacitance and conductance) to less than 2.23 will greatly increase the chances of having high average yield. Ensuring that the average of variable 21 (a capacitance measure) is above 468.9 eliminates further possibility for low yields. The average yields improve as the far-left branch of the tree is followed. Of course, if the standard deviation of variable 11 cannot be reduced to 2.23, then keeping it below 4.97 is the next best course of action.

V. Summary

We have illustrated the use of several multivariate analysis and modeling tools to study wafer probe test data. Although clustering analysis seems ideal for this particular problem, the results were disappointing. Results for principal component analysis were very difficult to interpret and did not represent as much of the inherent variability in the data as we had hoped. When the SAS procedure GENMOD was run on the principal components, the model was significant, but again, we faced the problem of interpreting the principal components. Using linear regression, we were able to obtain some interpretable results, but with great time and effort and violation of normal

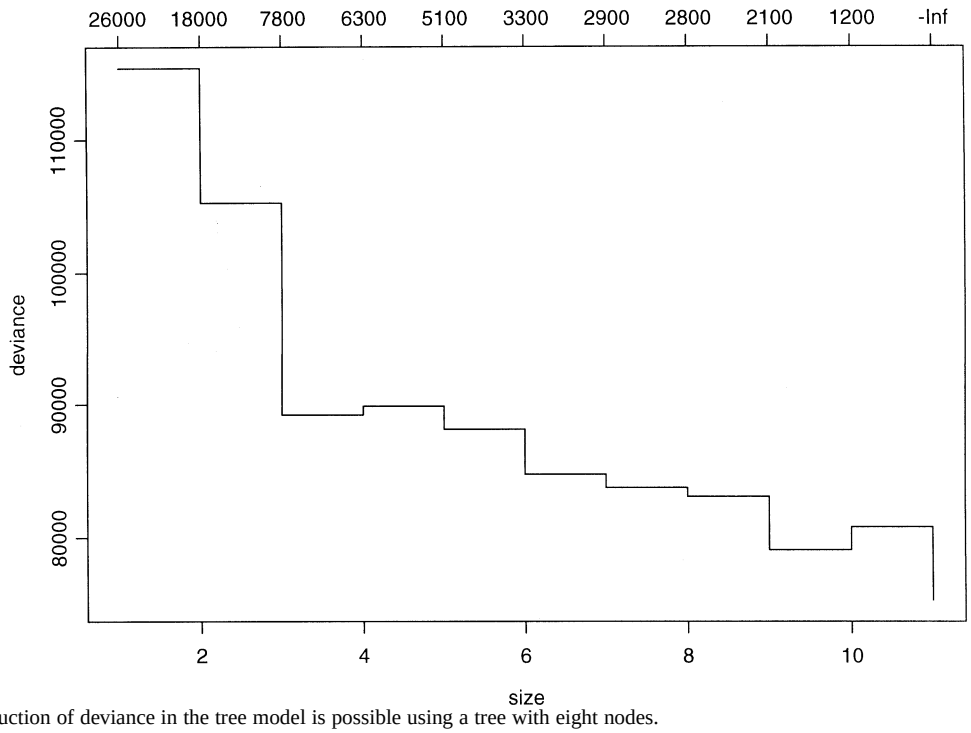


Fig. 6. Adequate reduction of deviance in the tree model is possible using a tree with eight nodes.

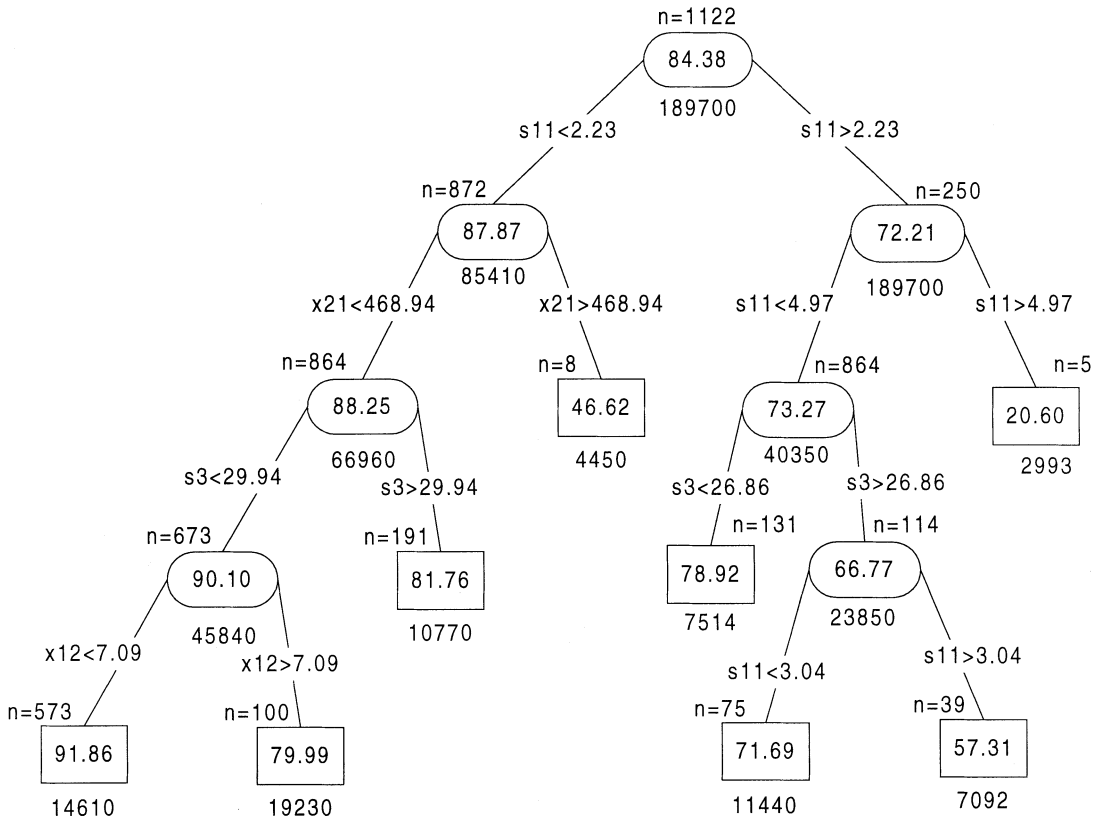


Fig. 7. Final tree after pruning to eight nodes. Ovals denote nonterminal nodes. Rectangles denote terminal nodes. Deviance values are beneath the corresponding node. Sample sizes at each node are denoted by "n."

regression assumptions. The splitting of the data according to the size of the residuals and to the yield indicated that the procedure of CART would be ideal.

The CART analysis proved useful. Both the problem and the data indicated a need for partitioning into high- and low-yield branches. CART is essentially a distribution-free procedure so the nonnormal yield data does not limit the application of the procedure. This provided results without the difficulties ordinarily associated with nonnormal data. The final tree after pruning produced a best path to high yield (note the path with

average yield equal to 91.9%) and a few paths to avoid (note the paths resulting in yields of 47%, 57%, and 21%). We strongly recommend CART as an exploratory and analysis technique for wafer probe data.

REFERENCES

[1] D. Boning and J. Chung, "Statistical metrology: Tools for understanding variation," *Future Fab Int.*, vol. 2, pp. 323–327, 1996.

[2] L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[3] C. Y. Chao and L. S. Milor, "Performance modeling using additive regression splines," *IEEE Trans. Computer-Aided Design*, vol. 6, pp. 1013–1022, 1987.

[4] V. Czitrom and P. D. Spagon, *Statistical Case Studies for Industrial Process Improvement*. Philadelphia, PA: SIAM, 1997.

[5] R. Gananadesikan, "Graphical methods for internal comparisons in ANOVA and MANOVA," in *Handbook of Statistics, Vol. 1: Analysis of Variance*, R. R. Krishnaiah, Ed. Amsterdam, The Netherlands: North Holland, 1980, pp. 133–177.

[6] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1992.

[7] R. Khattree and D. N. Naik, *Applied Multivariate Statistics With SAS® Software*. Cary, NC: SAS Institute, 1995.

[8] B. Koppenhoefer, S. Wuerthner, L. Ludwig, W. Rosenstiel, H.-H. Kuge, M. Hummel, and P. Federl, "Analysis of electrical test data using a neural network approach," in *Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf.*, 1997.

[9] *Minitab User's Guide*: Minitab Inc., 1998, Release 11.

[10] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley, 2001.

[11] D. F. Morrison, *Multivariate Statistical Methods*, 3rd ed. New York: McGraw-Hill, 1990.

[12] R. H. Myers, D. C. Montgomery, and G. G. Vining, *Generalized Linear Models With Applications in Engineering and the Sciences*. New York: Wiley, 2002.

[13] R. H. Myers and D. C. Montgomery, "A tutorial on generalized linear models," *J. Quality Technol.*, vol. 29, no. 3, pp. 274–291, 1997.

[14] *SAS Procedures Guide*. Cary, NC: SAS Institute Inc., 1997, version 6.0.

[15] *S-PLUS User's Guide*. Seattle, WA: MathSoft, Data Analysis Products Division, 1997, version 4.0.

[16] V. Sharma, "A new modeling methodology combining engineering and statistical modeling methods: A semiconductor manufacturing application (process control)," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, 1996.

[17] W. N. Venables and B. D. Ripley, *Modern Applied Statistics With S-PLUS*, 3rd ed. New York: Springer-Verlag, 1999.

Katina R. Skinner is currently working toward the Ph.D. degree in industrial engineering at Arizona State University, Tempe.

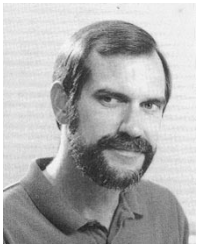
Her research interests include multivariate SPC and applied statistics.

Douglas C. Montgomery is Professor of Engineering at Arizona State University, Tempe. He is the author of several widely used texts on industrial statistics and over 130 technical papers.

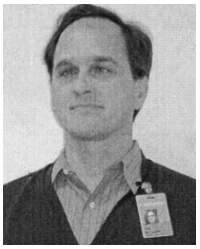
Prof. Montgomery is a Fellow of the American Statistical Association, the Royal Statistical Society, the American Society for Quality, and the Institute of Industrial Engineers. He is a former Editor of the *Journal of Quality Technology* and a current Chief Editor of *Quality and Reliability Engineering International*. He has received numerous awards for his research, including the Shewhart Medal of the ASQ.

George C. Runger holds degrees in industrial engineering and statistics.

He is a Professor in the Industrial Engineering Department at Arizona State University. He has coauthored research papers on design of experiments, real-time monitoring and control, data mining, and other data-analysis methods. He teaches these and related subjects in graduate courses at Arizona State University, Tempe. Currently, his work is funded by several grants from the National Science Foundation. He is a coauthor of *Applied Statistics and Probability for Engineers* (New York: Wiley, 2001, 2nd ed.) and *Engineering Statistics* (2001, 2nd ed.). In addition to academic and consulting work, he was a Senior Engineer at IBM.



John W. Fowler is an Associate Professor in the Industrial Engineering Department at Arizona State University (ASU), Tempe. Prior to his current position, he was a Senior Member of Technical Staff in the Modeling, CAD, and Statistical Methods Division of SEMATECH. His research interests include modeling, analysis, and control of semiconductor manufacturing systems. He is the co-director of the Modeling and Analysis of Semiconductor Manufacturing Laboratory at ASU. The lab has had research contracts with NSF, SRC, SEMATECH, Infineon Technologies, Intel, Motorola, ST Microelectronics, and Tefen, Ltd.

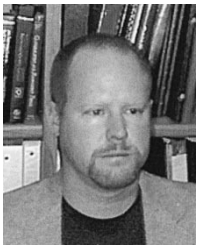


Daniel R. McCarville received the B.S. degree in mechanical engineering from the University of Arizona and the M.S. and Ph.D. degrees in industrial engineering from Arizona State University, Tempe.

He is a Director of Quality at Mindspeed Technologies, Newport Beach, CA. His research interests are in industrial statistics with emphasis on process control and optimization.

Teri Reed Rhoads received the B.S. degree in petroleum engineering from the University of Oklahoma, Tulsa, in 1985, the M.B.A. degree from the University of Texas of the Permian Basin in 1992, and the Ph.D. degree in industrial engineering from Arizona State University, Tempe, in 1999.

She is an Assistant Professor of Industrial Engineering at the University of Oklahoma. She also serves as the Assessment Advisor to the College of Engineering. She is actively involved in research with industry as well as with the National Science Foundation and the Department of Education. Her areas of interest are quality engineering as well as engineering education and assessment.



James D. Stanley received the B.S. degree in mathematics and the M.N.S. degree in statistics from Arizona State University, Tempe, in 1992 and 1994, respectively.

He is a Test Engineer and Statistician for the Sensor Products Division at Motorola, Tempe, AZ. He has published multiple articles on the application of statistical methods to semiconductor manufacturing.