U.S. Department of Commerce
Technology Administration
National Institute of Standards and Technology
Information Technology Laboratory

REPORT OF ACTIVITIES OF THE STATISTICAL ENGINEERING DIVISION

MARCH 1998

Covering Period: January 1997 – January 1998

Publications: January 1996 – January 1998

The cover displays the results of a clustering analysis of a large database of mass spectra. Each spectrum consists of approximately 150 values. The number of values was reduced to five summary values through the use of principal components. A k-means clustering procedure was then performed on these values. The figure displays all pairwise scatter plots of the five summary values with the color indicating cluster membership. The analysis will aid in the matching of spectra of unidentified compounds to a compounds in the database.

Contents

1. Division Overview

Keith R. Eberhardt, Acting Chief
Statistical Engineering Division, ITL

The Statistical Engineering Division (SED) is a unit of the Information Technology Laboratory of the National Institute of Standards and Technology (NIST). SED collaborates in NIST measurement science and technology research programs to support US industry through design of experiments, statistical modeling, and analysis and interpretation of data. We participate in the Laboratory's interdisciplinary research and development teams to advance information technology; we contribute to the development of appropriate statistical methodology, building on a foundation of pertinent topics in probability and mathematical statistics; and we provide leadership and computational tools to facilitate the implementation of modern statistical design, analysis and process control procedures.

The Division operates from both the Gaithersburg, Maryland, and the Boulder, Colorado, campuses of NIST with professional staff composed of Ph.D. and Masters degreed mathematical statisticians, of whom 16 are assigned to the Gaithersburg site and 3 are assigned to Boulder. This full-time staff is augmented by several faculty appointees, guest researchers and post doctoral students. A staff listing appears in Section 2.

Division priorities are driven by the need to support the NIST mission in the areas of:

- Promoting improved use of information technology through the NIST laboratories and outreach to industrial partners,
- Engaging in fundamental research in measurement sciences,
- Facilitating the Calibration and Standard Reference Materials programs, and
- Collaborating in high visibility projects of national interest.

This report provides technical summaries of some key project activities from January, 1997 to January, 1998 and a compilation of staff activities during that time. The project summaries are intended to provide a representative sampling, but not a comprehensive summary, of all Division activities. It is important that readers understand that all of this work is done collaboratively with other scientists and engineers, not by statisticians alone.

Indeed, there are many activities of SED that cannot be represented here. Additional information can be found on the SED World Wide Web Home Page by accessing the following URL: http://www.nist.gov/itl/div898/.

Thank you for reading. We welcome your comments. Please address them to:

Keith R. Eberhardt
Acting Chief, Statistical Engineering Division
Building 820, Room 353
National Institute of Standards and Technology
Gaithersburg, MD 20899
Email: Keith.Eberhardt@nist.gov

2. Staff

Division Staff

Keith Eberhardt, Acting Chief
Stephany Bailey, Secretary
Harry Ku, Intermittent
Robert Lundegard, Intermittent
Joan Rosenblatt, Guest Researcher
Andrew Rukhin, Faculty (U. of Maryland, Baltimore County)
Yudaya Sivathanu, Guest Researcher (Purdue U.)
Grace Yang, Faculty (U. of Maryland)

Raghu Kacker, Project Leader
David Banks
Charles Hagwood
James Yen

Measurement Process Evaluation Group

Carroll Croarkin, Group Leader
Mary Clark*, Secretary
Shirley Bremer, Intermittent
James Filliben
Lisa Gill
William Guthrie
Alan Heckert
Niels Hoven (Student)
Mark Reeder (Student)
Nien Fan Zhang

Statistical Modeling and Analysis Group

Keith Eberhardt, Group Leader
Eric Lagergren
Stefan Leigh
Mark Levenson
Hung-kung Liu
Mark Vangel

BOULDER
Dominic Vecchia, Project Leader
Lorna Buhse, Secretary
Duane Boes, Faculty (Colorado St. U.)
Kevin Coakley
Hariharan Iyer, Faculty (Colorado St. U.)
C.M. Jack Wang

**Part-time*

3. Project Summaries

3.1. Research in Information Technology

3.1.1. Testing Random Number Generators for Crypto-Security

David L. Banks, Stefan Leigh, Mark Levenson, Andrew Rukhin, Mark Vangel
Statistical Engineering Division, ITL

James Dray, James Nechvatal, Miles Smid, Juan Soto
Computer Security Division, ITL

Secure network communication depends on encryption technology. And all modern encryption technology requires the use of random number generators that can produce essentially random binary sequences. This project develops statistical tests for nonrandomness that reveal insecure generators.

Many generators have been proposed by the mathematical community. Some are known to be flawed, but have compensating advantages of speed or low memory requirements. Others are thought to be truly random (in a technical sense, that relates the unpatternedness of their behavior to the hardness of intransigent mathematical problems). But before any of these generators are used, their output should be subjected to a battery of statistical tests, to verify an acceptable level of pseudorandomness.

Many statistical tests exist that determine specific kinds of nonrandomness. These include: (1) Freqency tests, to verify that the proportions of zeroes and ones are equal; (2) Runs tests, to verify that the binary sequence doensn't alternate too frequently, or contain too many or too few long substrings of zeroes or ones; (3) Spectral tests, that look for periodicity in the local frequencies of zeroes and ones (an extension of this test based on wavelet analysis looks for aperiodic variation in these frequencies); (4) Geometric tests, which can look at how often a random walk based on the sequence returns to the origin, or at the local dimensionality of n-tuplets of numbers built from the sequence, and other theoretical properties of multivariate spacings; (5) Compression tests, that use complexity theory to estimate if a sequence compresses too much, indicating that some kind of patterning is present. We have developed a large suite of these tests, together with the associated mathematical theory, critical values, and guidelines for application, so that any user will be able to assess the strength of their own encryption algorithms.

Two issues are independence and coverage of the tests. Independence problems arise because some tests give similar results for nearly all sequences, and are redundant. We solve this using latent variable theory and principal components analysis.

The coverage problem arises because we want to ensure that our tests are able to capture the principle kinds of non-randomness that arise from current and future generators. (Matin-Löf proved that no finite set of tests can ever completely validate a generator, and thus the best possible result is to detect shortcomings that arise in practice.) To address this problem, we use the theory of simultaneous inference and a modified form of the Inclusion-Exclusion Principle.

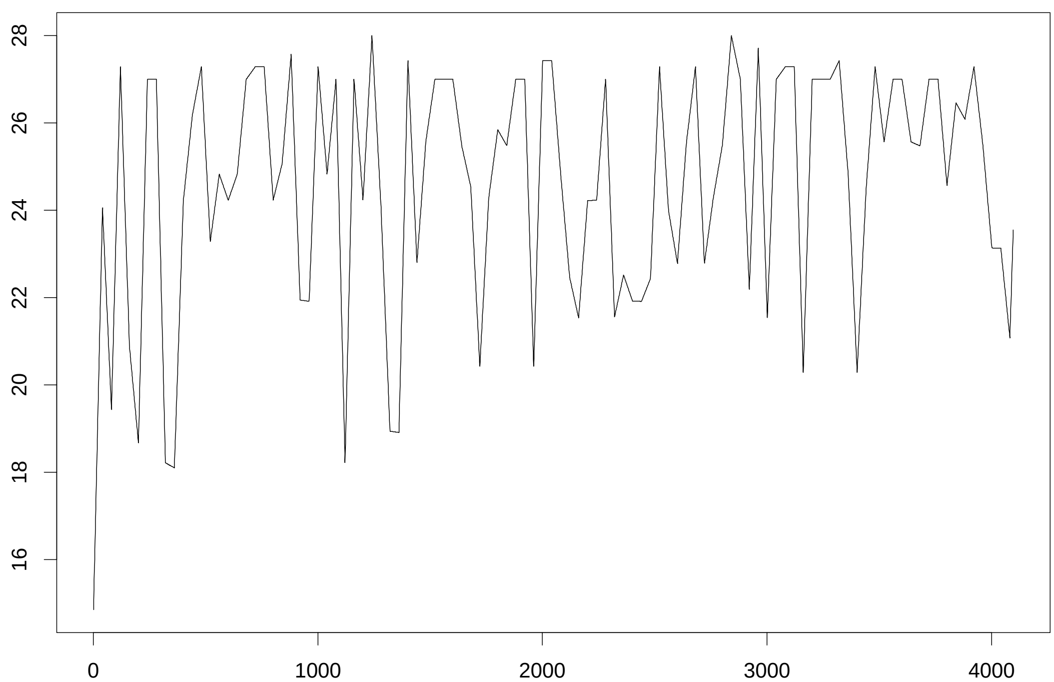


Figure 1: This figure shows the Lempel-Ziv complexity of the 4096 binary sequences of length twelve, from $0, \ldots, 0$ to $1, \ldots, 1$. The data have been smoothed by loess, and illustrate the complexity of the sequences by the length of their compression.

3.1.2. Statistical Methods for Software Validation

David L. Banks, Charles Hagwood, Raghu Kacker, James Yen
Statistical Engineering Division, ITL

Will Dashiell, Leonard Gallagher, Lynne Rosenthal
Software Diagnostics and Conformance Testing Division, ITL

Software reliability is the central problem of the Information Revolution. Faulty software can cost fortunes and lives; also, the worldwide effort to validate commercial code ties up an enormous investment of high-level human resources. The scale of this problem means that even modest statistical progress to reduce the testing burden while maintaining current performance levels will enable great technological advance.

Recently, new statistical methods have been proposed for conformance testing; these reduce costs, quantify uncertainty, or both. Our research program compares these methods, invents better ones, and determines which cocktail of techniques is most useful for specific classes of problems. We also evaluate the benefit that best-practice methods can confer, to support management assessment of software costs, and risks.

Currently, we have focused upon the four most promising directions in recent software validation: (1) Coverage designs (Dalal and Mallows, 1997), that allocate testing effort across modules or functions so as to ensure the joint exercise of all subsets of fixed size; (2) Usage models (Trammell, 1996), which attempt to allocate test effort according to usage data or mission criticality; (3) Optimal stoppage (Dalal and Mallows, 1989), which uses information on the catch times and severity of bugs to set up a dynamic programming problem whose solution (under some assumptions) determines the best time to release the code; (4) Extensions of binomial models for failures in a fixed test suite (cf. Sahinoglu and Spafford, 1990), using Bayesian and finite-population techniques. A key goal is comparative evaluation of these different approaches.

We have begun a simulation experiment that aims at giving first-order information on comparative performance of inspection protocols for virtual software. The experiment takes account of differential catchability among bugs, different costs of bugs (in terms of the damage caused if the undetected bug were released), different locations of the bugs (either in terms of module or function), and the fact that some kinds of bugs, such as logical errors, show cluster structure. This work is related to optimal search theory. We believe that simulated software tests have the potential to remove the greatest single barrier to progress in the evaluation of conformance testing methods.

Also, we are exploring the of clinical trials methods for deciding whether to release software. The decision to market a new drug is analogous to the decision to release new code—drug approval uses a four-step protocol established by the FDA. Software manufactures employ four similar steps, but are less systematic in combining the information at each stage. By transferring statistical techniques developed for drug testing, we hope to improve software testing.

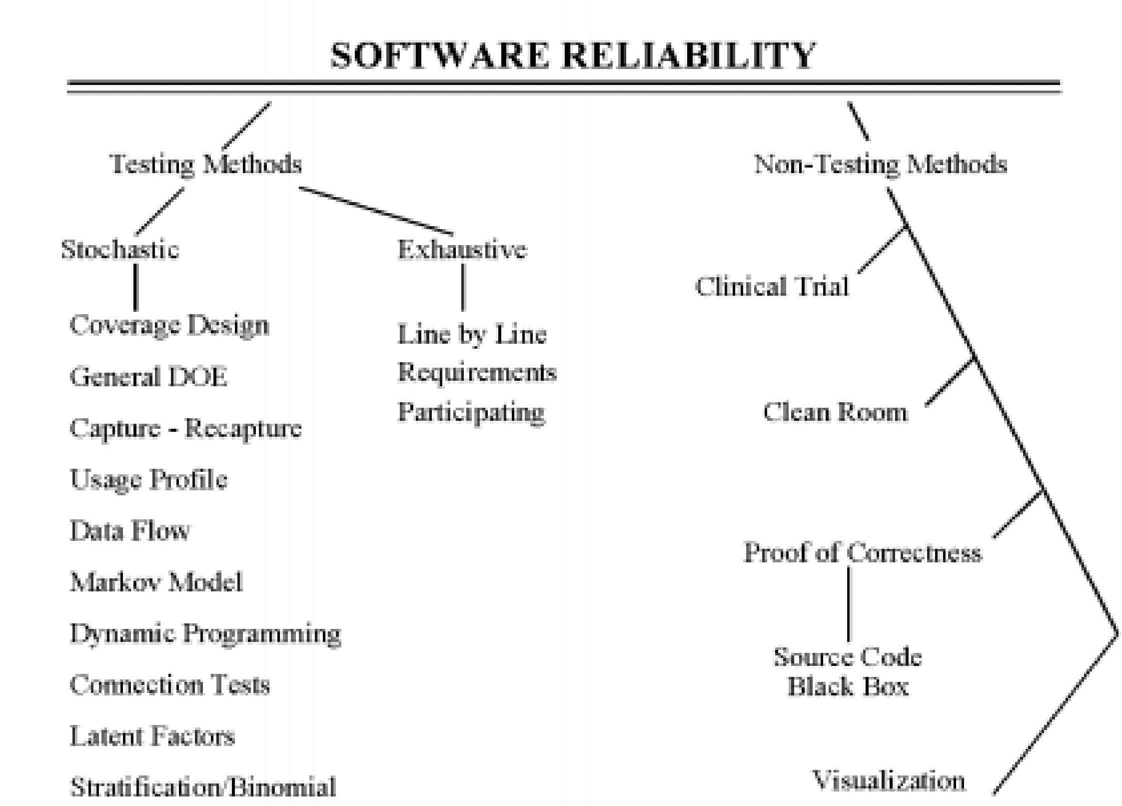


Figure 2: There are many strategies for software validation. This partial taxonomy organizes the major lines of research on this problem; however, because the field is large and methods may be applied in combination, this isn't a comprehensive representation.

3.1.3. Statistical Visualization for Managing Network Intrusion and Anomaly Detection

David L. Banks, Mark Levenson
Statistical Engineering Division, ITL

John Cugini, Joseph Konczal, Sharon Laskowski
Information Access and User Interfaces Division, ITL

Donald Marks
Computer Security Division, ITL

Experts in computer security are concerned with the need for intrusion/anomaly detection, and a number of automated system monitors are under development. Most of this research has not yet been informed by the results of NIST's work in the Information Exploration Shoot-Out, which used system audit logs with known intrusions as the testbed for a comparison of visualization-based data-mining methods.

Our project is to develop an interactive visualization tool that acts as a post-processor for the output of an automated system monitor, thereby presenting the system manager with information on threat levels and system anomalies in a more directly interpretable way. Also, the tool will enable managers to take direct action in response to a threat, or access information useful in evaluating an apparent threat. The Information Security Systems Company (ISS) has provided support for this research, as has DARPA and Roy Maxion's Harbinger Project at the School of Computer Science, Carnegie Mellon University.

Our interactive visualization tool is called NAIVE, for Network Anomaly/Intrusion Visualization and Exploration. The tool has two versions; one is appropriate for maintaining security in small local area networks, and the other is scalable to address the needs of very large systems. NAIVE is being designed to ensure platform-independence. The development platform is a Unix workstation; software is written using Tcl/Tk, a portable user-interface environment, and the visualization is being accomplished with OpenGL, a portable graphics library.

An example prototype has been built using DARPA's Common Detection Intrusion Framework. It presents real-time intrusion data visually as a graph. Nodes represent terminals and whose edges represent links between terminals. The edges carry color-coded information about the transmission of secure files, atypical usage patterns, and other performance anomalies. By clicking on an edge, the display can indicate which ports are involved in the transmission, and other detailed information.

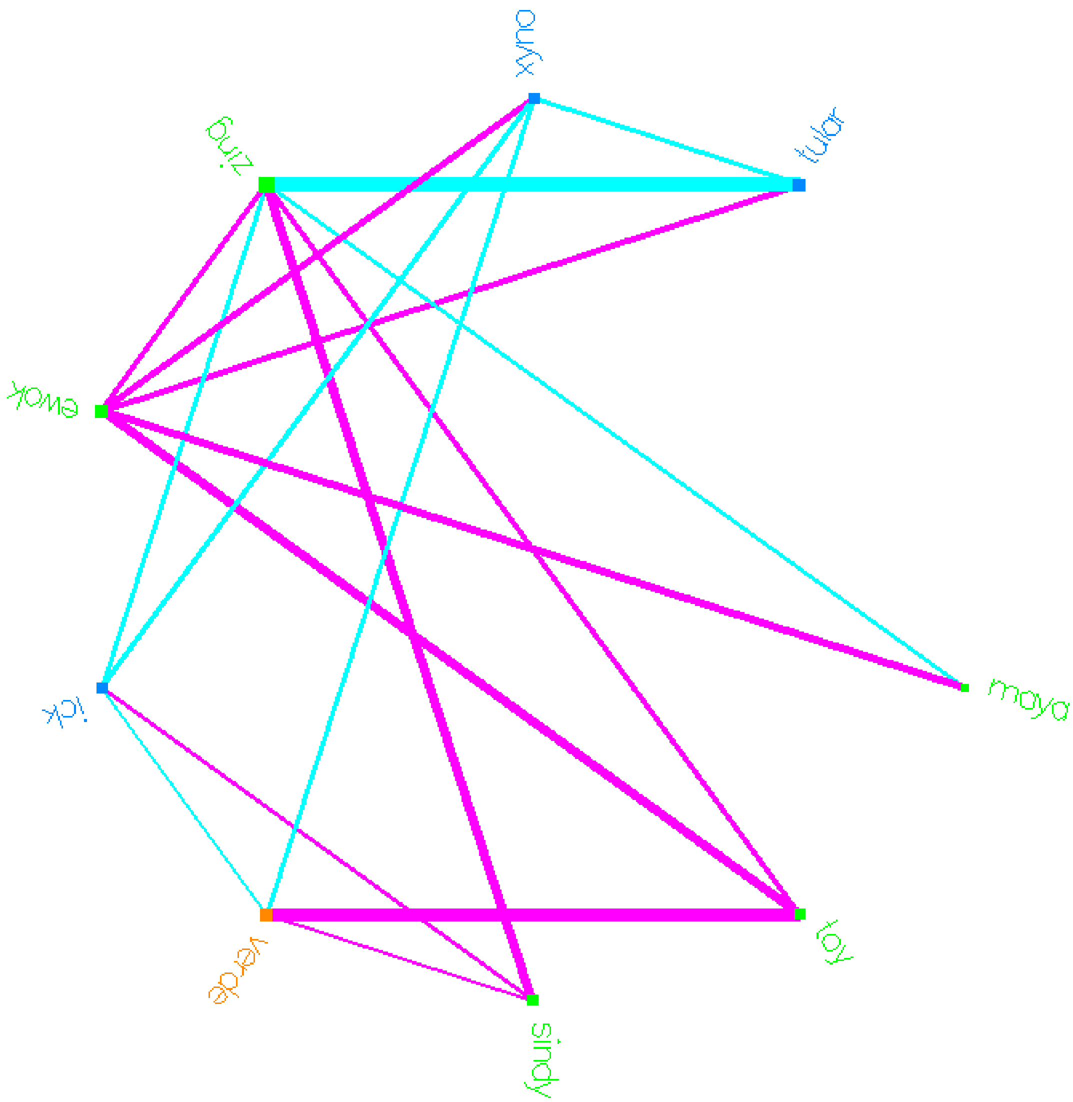


Figure 3: This figure shows a screendump of the NAIVE system, using simulated data to show threat levels and traffic on a small local area network. The colors of the edges show the level of threat, and their width indicates the volume of traffic. The colors of the nodes indicate categories of user.

3.1.4. Comparison of Information Retrieval Systems

David L. Banks, Eric Lagergren, Nien-Fan Zhang
Statistical Engineering Division, ITL

Donna Harman, Walter Liggett, Paul Over, Ellen Voorhees
Information Access and User Interfaces Division, ITL

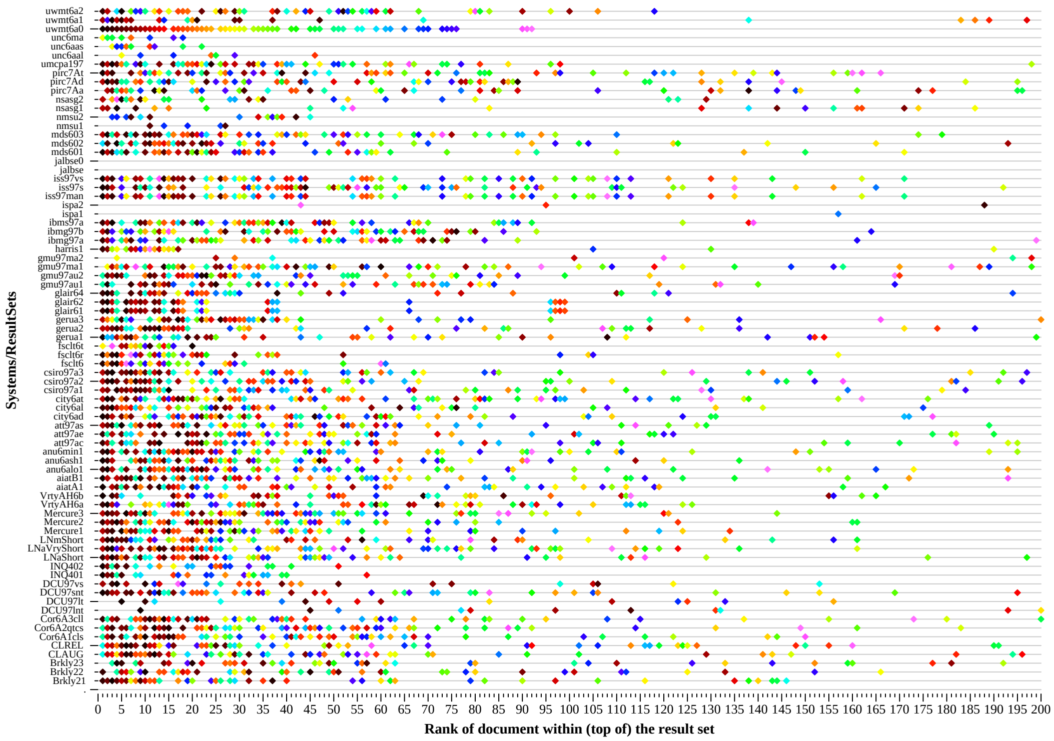
For many years, NIST has been a leader in text-retrieval research. This domain poses a number of statistical problems, and as we have built on previous progress, the remaining unsolved problems have grown increasingly more difficult. Now, the pressing issue is to compare the performances of different document-retrieval systems across a range of retrieval topics. A solution in this area will enable the many universities and industries that have designed document browsers to achieve a high-level understanding of the strengths and weaknesses of their products, and point the way to new improvements.

In the comparison problem, document-retrieval system i assigns a rank to document j for the kth retrieval topic; call this rank X_{ijk}. From this information, one wants to determine which topics are intrinsically hard or easy, and how those topics might be identified *a priori*. Following up on that, once one has some meaningful clusters of topics, one wants to find out which browsers work best with which clusters.

Our first examination used straightforward analysis of variance techniques, especially Mandel's 'bundle-of-lines' procedure to overcome the problem of testing for interaction when there is only one observation for each combination of factor levels. The results were definitive; there are interactive effects between topic and browser, and these effects are strongest with the most difficult topics.

This result suggested several analytical strategies, including block-cluster analysis, latent variable modelling (based on item response analysis in educational testing theory), and statistical models for rank-valued data. These analyses are at different stages of completion, but the emerging consensus is that none of them will enable a comprehensive solution to the broad problem.

We have developed an analysis based on a new graphical tool called the beadplot, shown on the following page. In the beadplot, the most relevant document is colored dark red, and less relevant documents are given colors that tend towards purple, according to the visible spectrum. Thus one can see in the beadplot how each system rates the documents in relevance (white spots correspond to irrelevant documents). Thus it is easy to identify groups of documents which tend to be retrieved together; some systems retrieve them early, and assign them low ranks, whereas others miss the documents entirely, or assign them higher ranks. Based on insights from these plots, we now represent each topic in a three-dimensional space whose axes reflect how easy it is to capture relevant documents, how many red herrings the document pool contains, and how much the systems differ in success for that topic.



Distribution of uwmt6a0's RELEVANT documents ranked 100 or better in other systems' result sets for topic 302

Figure 4: Beadplots show the rank at which each relevant document was retrieved by each of the text-retrieval systems. The rows correspond to the retrieval system, and the colored dots correspond to documents. Dots of the same color indicate the same document, and the order and spacings along the row indicate the ranks the documents were assigned.

3.1.5. Statistics of Software Conformance Testing

Charles Hagwood
Statistical Engineering Division, CAML

Lynne Rosenthal
Software Diagnostics and Conformance Testing Division, ITL

The reliability of a software product is the probability that it will function properly. By functioning properly we mean, the production of answers that are not deviant from that required. In discussing the statistical measurement of software reliability, it is convenient to think of two types of situations (i) those where the reliability measurement involves the occurrence of failure over time, i.e. the time until the next failure, and (ii) the static model, where the reliability depends only on the successful performance of software given an intensive input interrogation.

In the time dependent reliability model, the raw data available to the tester is a sequence $t_1, t_2, \ldots, t_n$ of execution times between successive failures. As failures occur, attempts are made to correct the underlying faults in the software. Because the errors are corrected when detected and the corrections do not lead to further errors, it is reasonable to assume that the number of failures occurring in the time interval $(0, t]$, N_t is a nonhomogeneous Poisson process. Thus the observed times can be regarded as realizations of random variables $T_1, T_2, \ldots, T_n$, the interarrival times of N_t. The problem is to compute the conditional distribution of T_k given that $S_{k-1} = \sum_{i=1}^{k-1} T_i = t$, and is given by

$$R(x \mid t) = P(T_k > x \mid S_{k-1} = t).$$

$R(x \mid t)$ is just the probability that a failure does not occur in $(t, t+x)$ given the last failure occurred at time t. Of course, $R(x \mid t)$ is calculated from the distributional properties of N_t and the observed failure times. Note that here the error detection times $t_1, t_2, \ldots, t_n$ will, more often than not, depend on the class of users.

In the static model or the time-independent model the software system is subjected to an intensive test suite in order to determine if it meets specification. Conformance testing captures the technical description of a specification and measures whether an implementation (i.e., software product or system) faithfully implements the specification. The goal is to provide some level of assurance that the requirements imposed by a specification are being met by implementations claiming conformance to that specification. Conformance testing can be applied to the broad spectrum of software domains and software specifications. Regardless of the domain or specification, conformance testing is black-box or functional testing. Specifically, the internal structure and behavior of the implementation is not considered in the testing process. The conformance suite of tests are derived solely from the specification.

Conformance test suites are designed by carefully choosing different input values, trying to design test cases that will invoke every functional requirement in the specification, at least once. One method of accomplishing this is to partition the input space into disjoint subdomains, E_i, where the E_i represent different homogeneous parts of the program that test similar aspects of the program. The statistics involved in this collaboration is to estimated $1 - p$, the probability that the software will not fail, based on this stratified sampling method. In most cases in conformance testing, the software is said to pass if and only if there are no failures. Given this scenario, the estimate of reliability based on classical methods produces a reliability of one, although a nontrivial confidence interval can be found. A Bayesian approach is used to derive a nontrivial estimate of the reliability. As examples we take several software implementations of the Computer Graphics Metafile (CGM), International Standard ISO/IEC 8632.

3.2. Promulgation of Measurement Standards

3.2.1. High Precision Mass Calibration

Carroll Croarkin
Statistical Engineering Division, ITL

Zeina Jabbour
Automated Production Technology Division, MEL

The purpose of this project is to re-design the weighing designs used for the highest precision mass measurements at NIST to take advantage of a newly acquired HK1000 balance.

Weighing designs are schemes for intercomparing reference standards with test weights in such a way that the least-squares values assigned to the test weights depend on known values of the reference standards. Mass calibrations at NIST start at the kilogram level with a 1,1,1,1 design that includes the two NIST reference kilograms, whose sum is the restraint for the system of equations, a test kilogram and a summation of weights totaling one kilogram.

Weights are calibrated in sets. A typical set consists of weights of 1kg, 500g, 200g, 200g, 100g, 50g, 20g, 20g, 10g, etc., in which case the summation in the first series is 500g+200g+200g+100g which becomes the restraint for the second series. The second design in the series consists of comparisons among 500g, 200g, 200g, 100g, 100g, and a summation of 50g+30g+20g weights and is referred to as a 5,2,2,1,1,1 design. The series of weighing designs continues in this manner down to the smallest weight.

The HK1000 balance is an enclosed automated weighing system with four pans which require equal loads. Therefore, a 5,2,2,1,1,1 design cannot be run on this balance. On the HK1000, all loads must be nominally equal, and we are restricted to four loads so that all designs are 1,1,1,1 designs.

The difficulty is to construct a series of designs for this balance that ties the values of all weights to the NIST kilograms at some desired level of uncertainty. The facing page shows a portion of a series for calibrating 200g and 100g weights from 2 sets of 5,2,2,1,1,1 weights where the value of the 200g+ 200g+ 200g+ 200g+ 100g+ 100g restraint comes from designs above this series.

A concern is that we may not achieve the level of precision that the HK1000 balance is capable of as we work through the entire series as: weights are placed on the pans; the balance is closed and allowed to come to thermal equilibrium; the system runs through the design at night; the balance is opened the next day; new weights are inserted; etc. It remains to be seen whether changes between weighings will significantly erode the inate capability of the balance.

OBS	2	2	2	2	1	1	1	1	1	1
Y(1)	+	−								
Y(2)	+		−							
Y(3)	+			−						
Y(4)		+	−							
Y(5)		+		−						
Y(6)			+	−						
Y(7)	+				−	−				
Y(8)	+						−	−		
Y(9)	+								−	−
Y(10)					+	+	−	−		
Y(11)					+	+			−	−
Y(12)							+	+	−	−
Y(13)		+			−	−				
Y(14)		+					−	−		
Y(15)		+							−	−
Y(16)					+	+	−	−		
Y(17)					+	+			−	−
Y(18)							+	+	−	−
Y(19)			+		−	−				
Y(20)			+				−	−		
Y(21)			+						−	−
Y(22)					+	+	−	−		
Y(23)					+	+			−	−
Y(24)							+	+	−	−
Y(25)				+	−	−				
Y(26)				+			−	−		
Y(27)				+					−	−
Y(28)					+	+	−	−		
Y(29)					+	+			−	−
Y(30)							+	+	−	−
Y(31)					+	−				
Y(32)					+		−			
Y(33)					+			−		
Y(34)						+	−			
Y(35)						+		−		
Y(36)							+	−		
Y(37)					+	−				
Y(38)					+				−	
Y(39)					+					−
Y(40)						+			−	
Y(41)						+				−
Y(42)									+	−
RESTRAINT	+	+	+	+	+	+				

3.2.2. MALDI Time-of-Flight Mass Spectrometry

James J. Filliben
Alan Heckert
Mark G. Vangel
Statistical Engineering Division, ITL

Charles Guttman
Polymers Division, MSEL

MALDI (Matrix-Assisted Laser Desorption Ionization) Time-of-Flight Mass Spectrometry is a newly-developed methodology which will serve as the basis for future NIST synthetic bio-polymeric SRMs. Such SRMs have as their output not a number, but a table of weights and integrals corresponding to the spectrometer's spectral output. This "signature" of the polymer will change depending on whether it is a homo-polymer or a co-polymer and which specific homo- or co-polymer is being analyzed.

MALDI is superior to the classic (moment-based) methods of light scattering, osmometry, and ultracentrifugation (which yield slightly better than 2% characterizations)–MALDI offers a characterization based not just on moments but on the entire distribution. Such distributional detail provides the analyst with greater insight, discrimination, and accuracy which in turn yields superior SRM values.

In practice, however, the MALDI time-of-flight mass spectrometer is not yet a finished "production" device–it is a state- of-the-art piece of equipment which has instrumental and procedural factors which have not yet been characterized, understood, or optimized–such must be done in order for the full force of this new technology to be brought on line to produce improved polymeric characterizations in general and higher-accuracy SRMs in particular.

Some of the more important factors which affect the quality of the MALDI spectrometer output are as follows: 1. size of crystal; 2. laser energy; 3. ion optics; 4. detector voltage; 5. accelerator voltage; 6. delayed extraction time; 7. delayed extraction voltage; 8. mode (linear or reflection); 9. number of shots; 10. evaporation rate; 11. polymeric concentration; 12. baseline calculations; and 13. scanning pattern.

It is expected that bringing the device on-line will take about a year. The early stages of the shake-down (testing for stability, drift, and repeatability) are currently underway. The plots on the following page reflect such early analyses– note the existence of unexpected spikes (at a frequency and harmonic) in the (statistical) spectrum.

Future collaboration will involve the construction and execution of orthogonal designed experiments to simultaneously and efficiently study the large number of factors so as to provide information about dominant factors, important interactions, best settings, and anomolous conditions.

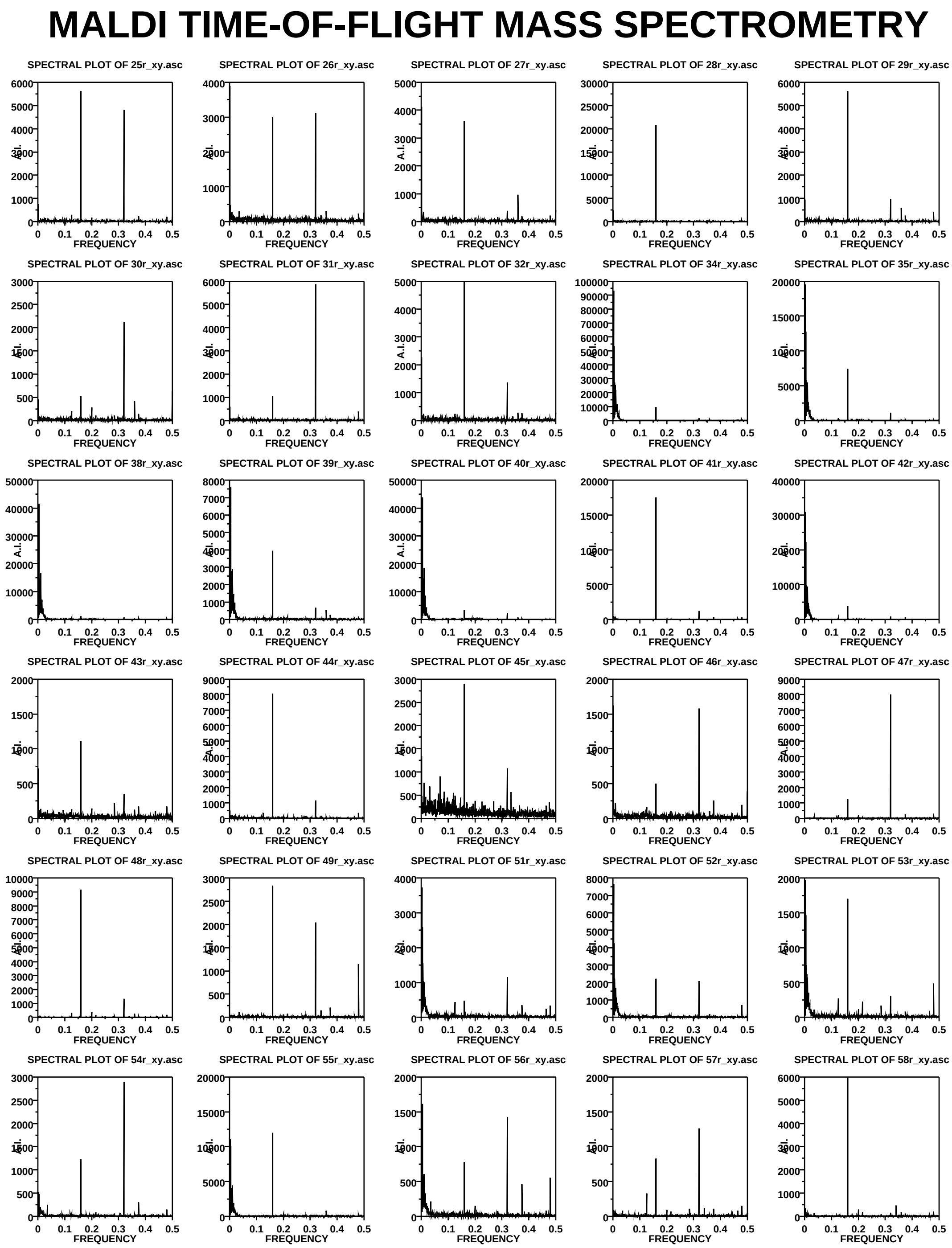


Figure 5: MALDI Time-of-Flight time traces.

3.2.3. Certification of Baby Food Composite, SRM 2383

Lisa M. Gill
Statistical Engineering Division, ITL

Katherine Sharpless
Analytical Chemistry Division, CSTL

In 1997, NIST released Standard Reference Material 2383, Baby Food Composite, which can be used as a control material when assigning values to in-house control materials and for validation of analytical methods for the measurements of proximates, vitamins, and minerals in baby foods and similar matrices. The "recipe" for SRM 2383 was developed at NIST. One thousand pounds of material were prepared by combining ingredients that go into commercially available jars of Gerber baby food; prepared creamed spinach and infant formula were added. The mixture was pumped into jars that held 70 g.

The certification of this SRM, which has 56 analytes, was broken up into three main categories: 1) Fat Soluble Vitamins and Cartenoids, 2) Water Soluble Vitamins, and 3) Proximates and Trace Minerals. The first category had measurements made by NIST and two outside laboratories. The other categories had data collected from a round-robin study and some additional participating laboratories. There was no data collected at NIST for these two areas.

The certification method used for the fat soluable vitamins and cartenoids has been applied to many chemical multimethod reference materials. For all the reporting methods, an equally weighted mean of the measurements is calculated with an expanded uncertainty at the 95% level of confidence with an additional bias allowance added on linearly. This method provides conservative results that have proven to be useful over time.

Due to the large number of participating laboratories for the water soluble vitamins, proximates and minerals, the certified value and associated uncertainty were derived by taking the mean of each of the lab means and then calculating the variance of the lab means. This method works well except when there are too few labs reporting data for a particular analyte; in these cases, the degrees of freedom is small, which in turn results in a large t-multiplier.

Three criteria were used for excluding data from the calculation of assigned values: 1) if for a particular analyte a laboratory's results did not agree with other laboratorys' results and that particular laboratory had poor quality control (QC) data; 2) data was not used if a laboratory did not provide QC data; and 3) outliers were determined based on the QC data and comparison with other laboratories utilizing the studentized residuals.

The uncertainty in the reference values is expressed as an expanded uncertainty according to the *ISO Guide to the Expression of Uncertainty in Measurement.*

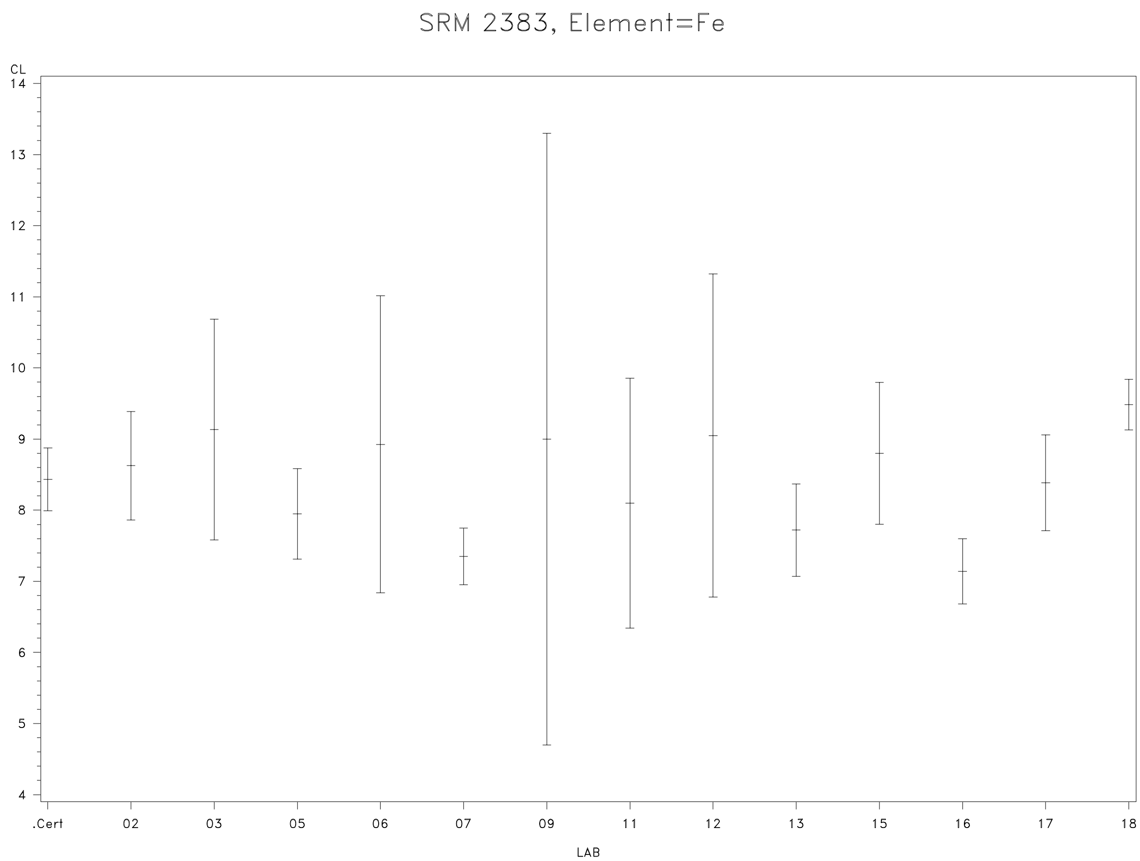


Figure 6: Comparison of the final assigned value (denoted .Cert) and its associated expanded uncertainty and the mean with expanded uncertainties for individual laboratory data for iron.

3.2.4. Estimation of Lithographic Overlay in Integrated Circuits Using Scanning Capacitance Microscopy

William F. Guthrie
Statistical Engineering Division, ITL

Santos Mayo, Joseph Kopanski
Semiconductor Electronics Division, EEEL

Misalignment of circuitry layers in IC's results in low yields and high costs. Thus, as feature sizes decrease, the ability to determine the amount of misalignment (called overlay) is becoming increasingly important. Overlay is typically measured using high-volume optical instruments. These instruments, however, are susceptible to systematic errors called tool-induced shifts (TIS). As a result, development of high-accuracy test methods for calibration of optical measurements is a priority for the semiconductor industry.

One promising test method is scanning capacitance microscopy (SCM). With this method, the probe of an atomic force microscope (AFM) is moved across a test structure on a wafer of IC's and changes in capacitance are recorded. Capacitance changes induced by different parts of the test structure can be used to estimate their locations.

Two test structures are shown in the adjacent figure (part A). For these structures the alignment of photoresist for a new layer of circuitry and lines from a previous layer are to be compared. The line in one structure is "floating" while the other is grounded. The floating line is simpler to make, but the grounded line gives a stronger signal.

SCM data from an electrostatic field simulator is shown in part B of the figure. Use of simulated data allows testing of the measurement system before fabrication of expensive samples for AFM use. The peaks near $x = 7.25$ and $x = 22.75$ correspond to the lines. The asymmetric peaks beside the line-induced peaks mark the edges of the photoresist.

Line locations are found by fitting a rational function in $(x - p_s)^2$, where p_s is the point of symmetry, to the line-induced peak data (part C). The points on either side of a first approximation to p_s are indicated by different symbols. Part D of the figure shows the peak data plotted versus the predictor variable based on the final estimate of p_s. The fitted function shows the quality of the fit in global terms. The locations of the photoresist openings are found by averaging the x values associated the minimum capactiance near each peak. Careful probe placement is key to the success of this technique.

The difference in line and photoresist locations gives overlay. Several examples are shown below. There is modest bias in the result with non-zero overlay for the grounded line and large bias for the floating line. Future work will focus on eliminating this bias, which stems from asymmetry in the data caused by overlap of photoresist and embedded lines.

	Floating Line	Grounded Line
True Overlay: 0.0000 μm	0.0021 ± 0.0031	-0.0010 ± 0.0030
True Overlay: -0.0500 μm	-0.0040 ± 0.0030	-0.0438 ± 0.0036

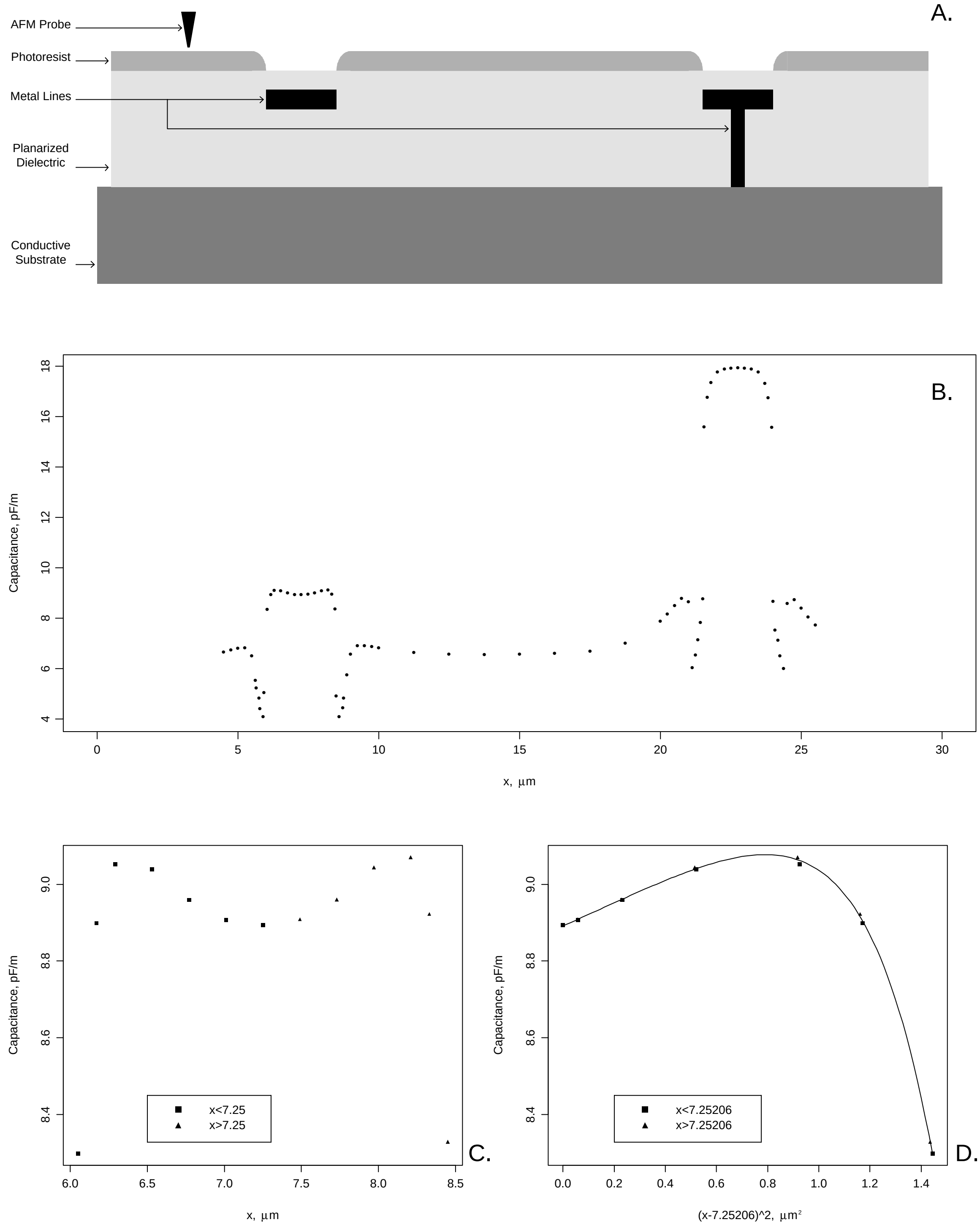


Figure 7: Test structures and data for measurement of lithographic overlay using scanning capacitance microscopy. A) shows the two test structures (floating and grounded lines) and the AFM probe. B) shows simulated SCM data from the two structures. C) shows a subset of the data used for estimation of peak location for the floating line. D) shows the data from C) "folded" around the peak location (the point of symmetry) and the fitted rational function.

3.2.5. On-line Monitoring of SRM's

Alan Heckert
Carroll Croarkin
Statistical Engineering Division, ITL

Tom Gills
Standard Reference Materials Program, TS

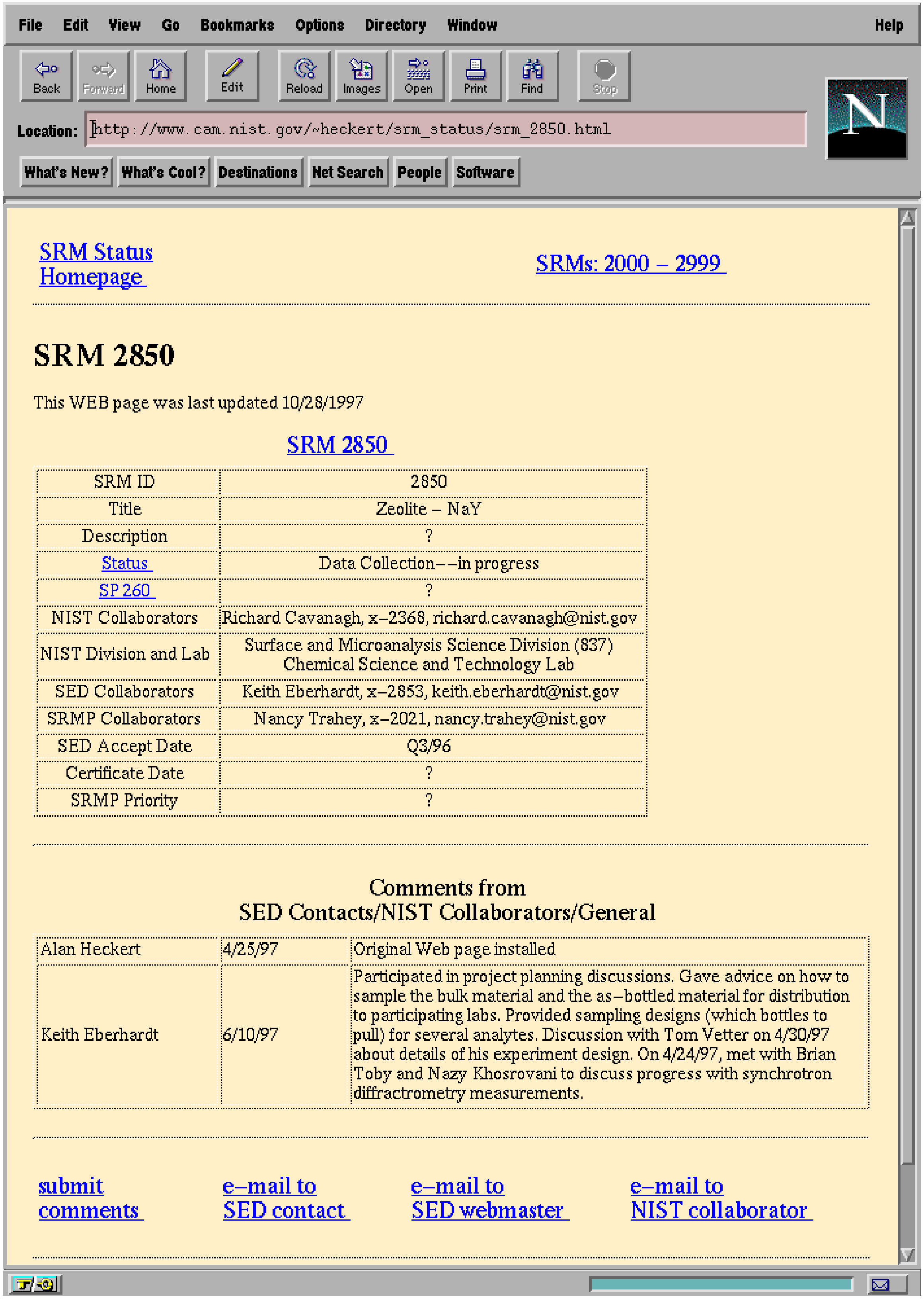
The Statistical Engineering Division is involved in the certification of a large number of Standard Reference Materials (SRM's). The certification of an SRM is a collaboration between one of the technical laboratories, the Office of Measurement Services, and the Statsitical Engineering Division. At the request of Tom Gills, Standard Reference Materials Program division chief, we investigated a method to provide on-line monitoring of the current status of all SRM's that SED is involved with. Previously, the status of the SRM's was maintained in a PC database with periodic hardcopy reports generated for the Office of Measurement Services.

Our primary goals in designing the on-line system were to have a system that all interested parties (the technical lab, the SRM office, and SED) could easily access, that would be easy to update, and that would be relatively low maintenance.

To achieve those goals, we created a web site accessible from the internal NIST web server (URL: http://www.i-nist.gov/itl/div898/index.html). The main home page provides 2 primary tables. The first provides access to SRM's via their number. This division is arbitrary and is provided simply to make the subsequent pages a more managable size. The second table provides access to the SRM's via the NIST laboratory. If the user clicks on the "Chemical Sciences and Technology" button, a summary table for all the SRM's in CSTL is generated. For convenience, this table is divided into the various CSTL divisions. Completed SRM's are listed separately from the currently active SRM's. The summary line lists the ID, name, contacts, and a status number. A similar scheme is followed for the other NIST labs and for the numerically ordered SRM's.

The user can click on any individual SRM to get a more detailed summary. The information for a given SRM can be updated from this more detailed page. Specifically, the current status can be modified and comment entries can be added. This updating is typically performed by the SED staff, but comment entries can be entered by any of the interested parties. New SRM's can be added from a form accessible from the home page. This updating is accomplished with Perl based CGI scripts.

Since this is a web based system, it easily accessible to any NIST staff member with a web browser. It is straightforward for SED staff to update the current status of their SRM's and to add new SRM's. Once the initial pages were created and the CGI scripts were written, minimal maintenance is involved.



SRM Status Homepage

SRMs: 2000 – 2999

SRM 2850

This WEB page was last updated 10/28/1997

SRM 2850

SRM ID	2850
Title	Zeolite – NaY
Description	?
Status	Data Collection––in progress
SP 260	?
NIST Collaborators	Richard Cavanagh, x–2368, richard.cavanagh@nist.gov
NIST Division and Lab	Surface and Microanalysis Science Division (837) Chemical Science and Technology Lab
SED Collaborators	Keith Eberhardt, x–2853, keith.eberhardt@nist.gov
SRMP Collaborators	Nancy Trahey, x–2021, nancy.trahey@nist.gov
SED Accept Date	Q3/96
Certificate Date	?
SRMP Priority	?

Comments from SED Contacts/NIST Collaborators/General

Alan Heckert	4/25/97	Original Web page installed
Keith Eberhardt	6/10/97	Participated in project planning discussions. Gave advice on how to sample the bulk material and the as–bottled material for distribution to participating labs. Provided sampling designs (which bottles to pull) for several analytes. Discussion with Tom Vetter on 4/30/97 about details of his experiment design. On 4/24/97, met with Brian Toby and Nazy Khosrovani to discuss progress with synchrotron diffractrometry measurements.

submit comments | e–mail to SED contact | e–mail to SED webmaster | e–mail to NIST collaborator

Figure 8: A sample page from the SRM web page.

3.2.6. Computational Metrology of Manufactured Parts

Mark Levenson, Keith Eberhardt
Statistical Engineering Division, ITL

Steven Phillips, W. Tyler Estler, Bruce Borchardt, Daniel Sawyer
Precision Engineering Division, MEL

Marjorie McClain
Mathematical and Computational Sciences Division, ITL

Yin-Lin Shen
Department of Mechanical Engineering, George Washington University

Coordinate Measuring Machines (CMMs) are used to measure the physical dimensions of manufactured parts. Their ability to measure almost an endless variety of geometrically complex parts in a rapid and accurate manner has led to their widespread use in industry. However, the sophistication and flexibility of the CMM make assessment of the measurement uncertainty difficult. Developing reliable uncertainty methodology for CMMs would (1) promote improvement in quality and efficiency through better determination of part dimensions and (2) facilitate international trade that requires ISO 9000 compliance.

Basically, a CMM is a robotic machine that positions a sensing probe in its working volume. The probe contacts a sample of locations on the part surface and the CMM records corresponding three-dimensional point coordinates. The measurement process contains many sources of uncertainty. Some of the largest sources are the geometric distortions of the machine frame, the systematic effect of the probe, and the thermal and mechanical effects of the operating environment. In the first two years of the project, our group developed a reliable model for real-time correction of the systematic effect of the probe. The result is an improved system without significant added costs. The papers "Error Compensation for CMM Touch Trigger Probes" and "Practical Aspects of Touch Trigger Probe Error Compensation" in *Precision Engineering* summarize the results. The latter paper includes a detailed uncertainty analysis.

Presently, based on three years of experience in the area, our group is working on several problems related to CMMs. First, we are continuing to quantify the major sources of uncertainty by examining sampling strategies to measure the geometric distortions of the machine frame. Second, we are developing a computer simulation model that produces uncertainty estimates on an arbitrary CMM measurement based on a small set of performance measures. Finally, using Bayesian methodology, we have derived superior decision rules for accepting or rejecting parts based on engineering tolerance and measurement uncertainty. The accompanying figure displays the results of a sensitivity analysis of the ISO standard versus the proposed Bayesian rule. In the figure, 100% corresponds to the correct estimate. Note that the ISO rule does not make use of prior uncertainty. Under various estimates of the uncertainties, the Bayesian rule results in lower cost than the ISO rule.

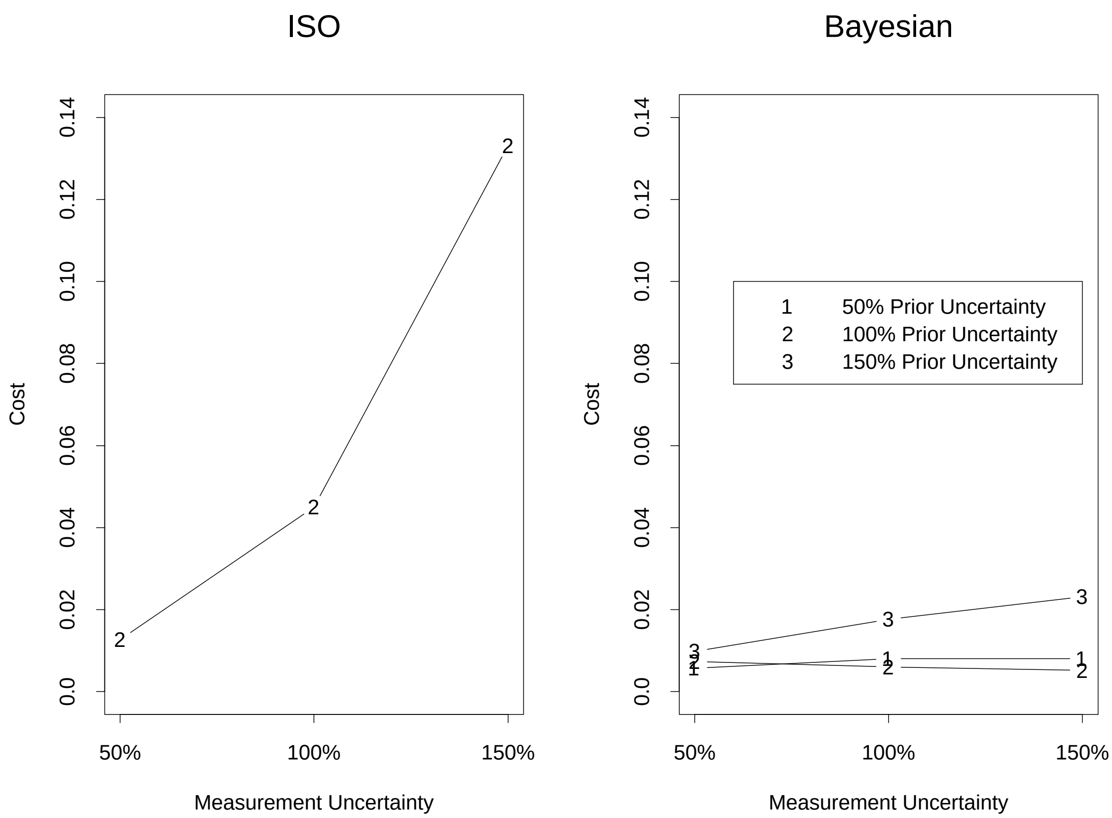


Figure 9: Costs of Part Acceptance Rules.

3.2.7. 1997 Interlaboratory Comparison of 10 V Josephson Voltage Standards

C.M. Wang
Statistical Engineering Division, ITL

Clark A. Hamilton
Electromagnetic Technology Division, EEEL

In the 1997 Josephson Voltage Standard (JVS) Interlaboratory Comparison (ILC), a traveling set of four 10 V Zener reference standards was circulated among the 15 participating standards laboratories. Each laboratory made 16 measurements of each of the four Zener standards over a period of 2–6 days. All of the laboratories used 10 V Josephson standards to measure the output voltage of the four standards. A comparison of the measurements of all laboratories yielded data on the uncertainty of each laboratory's measurement procedure, the performance of the traveling standards, and any offsets that may exist between the reference and testing laboratories.

In some previous JVS interlaboratory comparisons, one laboratory has served as a pivot with the traveling standards returning to the pivot after each set of measurements by a satellite laboratory. This is a big effort and in 1997 no laboratory volunteered to be the pivot laboratory. Thus, in this ILC, the standards traveled from one laboratory to the next until all 15 laboratories had participated. NIST–Boulder contributed three sets of measurements near the beginning, middle, and end of the ILC.

It is well known that Zener standards drift with time, and their voltage has a small and unknown dependence on pressure and other factors. A pivot laboratory allows the drift with time of the standard to be determined. Without a pivot laboratory, we used the following model to describe voltage measurements

$$V_{lsk} = V_{0s} + c_{1s}t_{lsk} + c_{2s}P_l + L_{ls} + \epsilon_{lsk},$$

where V_{lsk} represents the kth measured voltage for the sth standard of the lth laboratory, t_{lsk} is the elapsed time (in days), and P_l is the laboratory pressure (in kilopascals). In addition, $l = 1, 2, \cdots, 17$, $s = 1, 2, 3, 4$, and $k = 1, 2, \cdots, 16$. The quantity L_{ls} is the laboratory effect and assumed distributed as Gaussian with mean 0 and variance σ_{bs}^2. The random errors ϵ_{lsk} are all independent and assumed distributed as Gaussian with mean 0 and variance σ_{ls}^2.

Maximum likelihood was used to estimate σ_{bs}^2 and σ_{ls}^2 after adjusting for the drift and pressure effects. Least–median–of–squares regression was used to identify multiple outliers. The between– and the within–laboratory standard deviations, averaging over the four standards, were found to be 80 nV and 176 nV, respectively. The uncertainty of the ILC is obtained by combining the between– and the within–laboratory variations for the 64 measurements made at each laboratory. Specifically, the uncertainty (1σ) is $\sqrt{80^2 + 176^2/64} = 83$ nV, or 8.3 parts in 10^9.

This work will appear in the February or May 1998 issue of *Metrologia*.

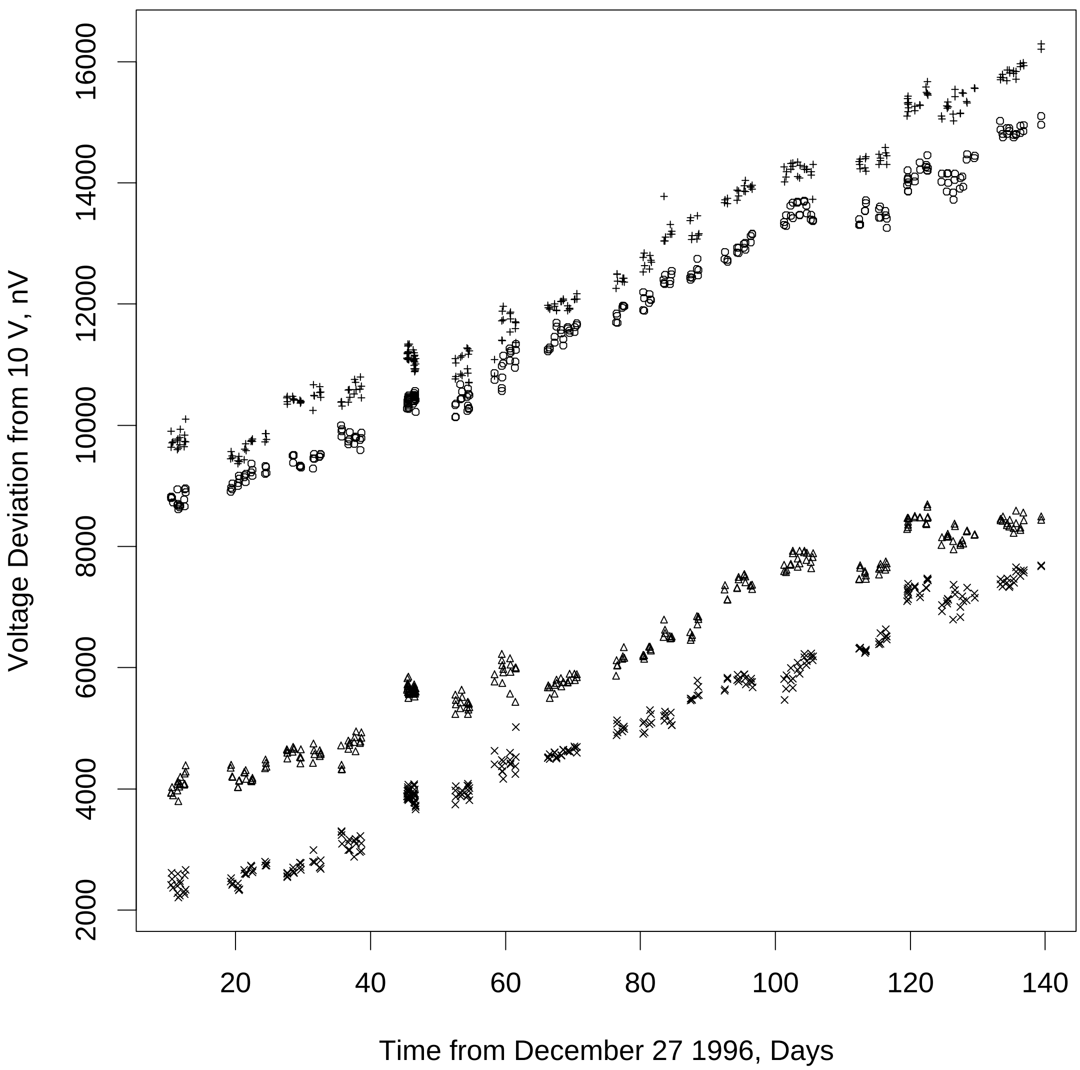


Figure 10: Voltage measurements of all participants. The horizontal and vertical axes are, respectively, the time in days from December 27, 1996 and the deviation from 10 V (in nV) of the measured voltage. The plotting symbols "○", "△", "+", and "×" are for standards 1, 2, 3, and 4, respectively.

3.2.8. Measurements of Polarization Mode Dispersion in Optical Fibers

C.M. Wang
Statistical Engineering Division, ITL

Paul A. Williams
Optoelectronics Division, EEEL

Polarization dispersion arises in single–mode fibers when there is imperfect circular symmetry in the fiber core. An optical pulse input to a fiber is split into two orthogonally polarized pulses. Distortion arises as a result of a differential group–delay (DGD) time between these two pulses at the output. Since DGD can have a limiting effect on the speed of digital communication systems and therefore is a good indicator of the performance of a lightwave system, it is routinely measured both at the manufacturing stage and in installed systems.

In long–fiber spans, DGD is a random effect, since it depends on the details of the birefringence along the entire fiber length. It is also sensitive to temperature and mechanical perturbations of the fiber. For this reason, a useful way to characterize DGD in long fibers is in terms of its expected value, or polarization mode dispersion (PMD).

Among the methods of PMD measurement, the fixed analyzer technique is perhaps the simplest to use. Light is polarized with an input polarizer and then launched into the test fiber. The transmission, through an output analyzer, is measured as a function of frequency (or wavelength). Based on the normalized transmission spectrum, $T(\omega)$, PMD is estimated by either

$$\langle \Delta\tau \rangle = 4N_m/\Delta\omega$$

or

$$\langle \Delta\tau \rangle = 0.824\pi N_e/\Delta\omega$$

where $\Delta\omega$ is the width of the frequency window over which the measurement is taken, N_m is the number of mean–value (0.5) crossings, and N_e is the number of extrema in the frequency window $\Delta\omega$. The factor 4 was obtained analytically, while the factor 0.824 was obtained through simulation, in the literature. Under regular conditions, N_m and N_e depend on $\langle \Delta\tau \rangle$ and $\Delta\omega$ only through their product. That is, the smaller the PMD, the bigger the window size required in order to estimate the PMD with the same precision. Since $T(\omega)$ is everywhere differentiable and continuous in ω, the number of discrete frequency measurements made will affect the outcome of N_m and N_e.

Simulation was used to study the effects of sampling density on the estimation of PMD. Fibers were simulated as a stack of 2700 waveplates with their optic axes randomly oriented. A new value for the polarization mode coupling factor of 0.805 (a 2% discrepancy with the old value of 0.824) was found. Systematic biases due to sampling density were quantified, and a simple correction algorithm was proposed. This and related work will appear in the *Journal of Lightwave Technology*.

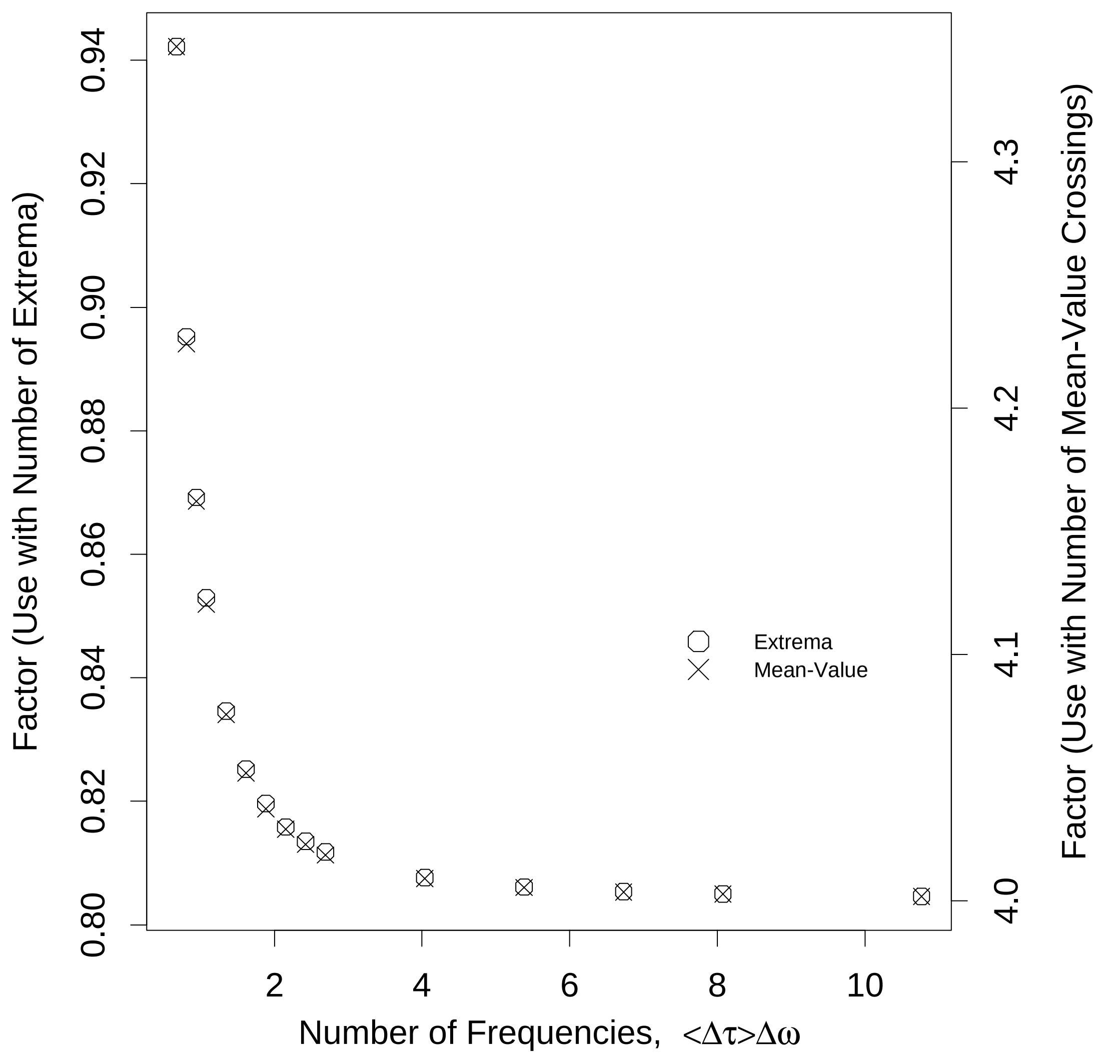


Figure 11: Calculated mode coupling constants versus sampling density. The horizontal axis is the value of $\eta = n_f/\langle\Delta\tau\rangle\Delta\omega$ where n_f is the number of points used to sample $T(\omega)$. The right and left vertical axes are, respectively, the mode coupling constants (k_1 and k_2) associated with N_m and N_e. As η becomes large, $k_1 \rightarrow 4$ and $k_2 \rightarrow 0.805$.

3.2.9. Estimating The Measurement of Pitch in Metrology Instruments

Nien Fan Zhang
Statistical Engineering Division, ITL

Michael T. Postek
Robert D. Larrabee
Precision Engineering Division, MEL

NIST is in a process of developing a new low-accelerating-voltage scanning electron microscope (SEM) magnification calibration reference standard 2090. This standard will be useful for all applications in which the SEM is currently being used, but it has been specially tailored for many of the particular needs of the semiconductor industry. In order for the NIST certification process to be complete, an estimate of the pitch measurement and its uncertainty must be evaluated. As the precision and accuracy of metrology instruments are pushed to the nanometer level, the evaluation of the performance of the pitch measurement algorithm becomes increasingly important. Figure 1 shows the diagram of the NIST SRM 2090a prototype SEM magnification standard. The left part is a lowest magnification drawing showing the 3 mm and 1 mm pitch patterns, while the right part of Figure 1 is a high magnification showing the two 4 micrometer (μm) and eight 0.2 μm pitch structures as well as the focusing and astigmatism-correction crosses.

The prototype SRM 2090a data was obtained by using the NIST SEM-based metrology system. A pitch distance between two pitch structures is defined as the distance between the left (or right) edge of one pitch structure and the left (or right) edge of another pitch structure. Mathematically, when the SEM signals at the edges are parallel straight lines the pitch distance is uniquely defined. However, in reality, when measurements are done by an SEM system as described above, the edges formed by discrete data points are not necessarily parallel.

Traditionally, a least squares regression line is fitted to the data points corresponding to each of the left (or right) edges of a pitch structure. Then, the distance between the two fitted lines (corresponding to two left or two right edges) at a certain height on the vertical axis is assigned as the pitch distance between the pitch structures. A disadvantage for this approach is that the pitch distance varies with the height at the vertical axis because in general these two fitted regression lines are not parallel. Another disadvantage for the traditional algorithm is that it is difficult to estimate the uncertainty of the pitch distance. We developed a statistical model based algorithm to eliminate this kind of uncertainty. The estimator of pitch distance and its uncertainty have been derived. Evaluations based on simulations show that the uncertainty of measurement of the pitch distance by the new method is smaller than that by the traditional one.

This paper has been published in *Metrologia* (1997), 34.

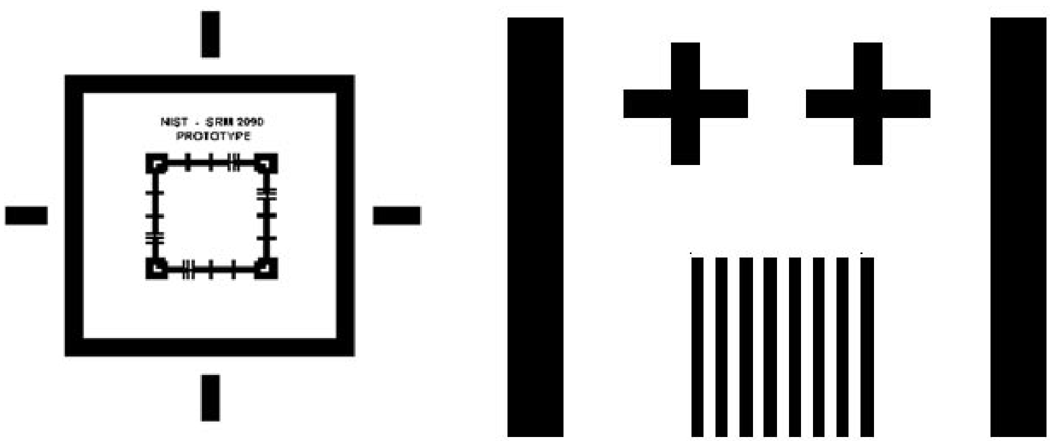


Figure 12: This figure shows the diagram of the SRM 2090a prototype SEM magnification standard.

3.2.10. A Statistical Measure for the Sharpness of SEM Images

Nien Fan Zhang
Statistical Engineering Division, ITL

Michael T. Postek
Robert D. Larrabee
Precision Engineering Division, MEL

Fully automated or semi-automated scanning electron microscopes (SEM) are now commonly used in semiconductor production and other forms of manufacturing. Testing and proving that the instrument is performing at a satisfactory level of sharpness is an important aspect of quality control. Industrial users of SEMs would like to have these instruments fuction without human intervention for a long periods of time, and to have some simple criterion (or indication) of when they need attention. At the present time, no self-testing is incorporated into these instruments to verify that the instrument is performing at a satisfactory performance level.

In this paper, a statistical measure, known as the multivariate kurtosis, is proposed as one approach to the measurement of the sharpness of SEM images. The application of Fourier analysis techniques to the SEM images is useful for sharpness measurement. The two-dimensional spectrum density of an image is similar to a probability density. In the theory of probability, kurtosis is a measure of a type of departure of a probablity distribution from the normal shape. The value of kurtosis can be compared with 3 to determine whether the distribution is "peaked" or "flat-topped" relative to a normal probability density. Literature shows that the smaller the kurtosis, the flatter the top of the distribution. The results have been extended to the multivariate case.

Based on the computed spatial frequency spectrum of selected SEM images, we observe that when an SEM image is visually sharper than a second image, the higher spatial frequency components of the first image are larger than those of the second. Treating the normalized spectrum as a probability density function, a sharper SEM image corresponds to a spectrum which has a larger shoulder or has a flatter shape. Thus, it can be concluded that the corresponding kurtosis of the sharper image is smaller. In addition to that, the marginal kurtosis can be used to measure the shape of the marginal spectrum. The difference between the marginal kurtoses, which are the kurtoses of the marginal distributions can be used to detect possible instrument vibration.

A series of five micrographs are selected as examples depicting a representative set of experiments to demonstrate the sharpness analysis procedure. The first figure is the graphical measure of sharpness following analysis of these five images. Low numbers for kurtosis indicate a better quality image or higher sharpness. Marginal kurtosis analysis also has been done. The second figure shows that something is wrong with Sample 4 because of the larger relative difference of marginal kurtosis. The results of kurtosis analysis coincide with the ranking of the quality of the samples.

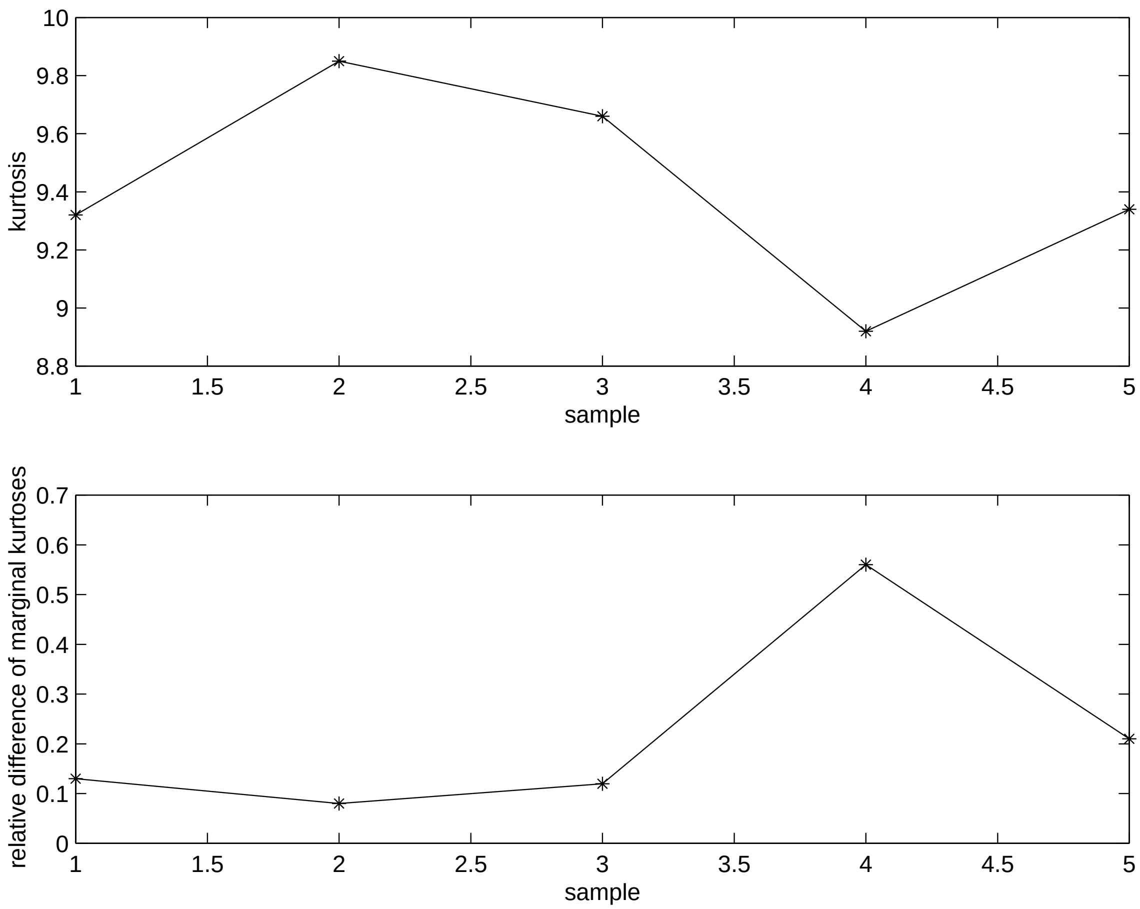


Figure 13: These figures show the sharpness measures for the five samples

This work was presented by N. F. Zhang at the SPIE's (The International Society for Optical Engineering) 1997 International Symposium on Microlithography and has been published in *Proceedings SPIE* 1997, vol 2050, p. 375-386.

3.3. Experiment Planning and Interpretation

3.3.1. Magnetic Trapping of Ultra Cold Neutrons and Determination of the Mean Lifetime of the Neutron

Kevin J. Coakley, Grace L. Yang
Statistical Engineering Division, ITL

M.S.Dewey, D.Gilliam
Ionizing Radiation Division, PL

In collaboration with researchers from Harvard University, Los Alamos National Laboratory, and University of Berlin, NIST plans to produce and confine polarized Ultra Cold Neutrons (UCN) in a magnetic trap. Based on this new technology, the neutron lifetime will be determined at a precision up to 100 times better than the current value. Along with other experimental data, a measurement of the mean lifetime of the neutron allows one to test the consistency of the standard model of electroweak interactions. Further, the mean lifetime of the neutron is an important parameter in astrophysical theories. Statistical and computational work has focused on optimal experimental design and dynamical studies of marginally trapped neutrons.

Optimal Estimation. There will be many run cycles of a two stage experiment. In the first stage of each run, neutrons from the NIST Cold Neutron Research Facility are guided into a superfluid ^{4}He bath where they dissipate almost all their energy by inelastic scattering. These UCN are confined in a magnetic trap. After filling the trap to some level, the neutron beam is blocked from entering the trap. During the decay stage of each run, decay events, as well as background events, are recorded. Denote the duration of each stage as T_{fill} and T_{decay}. Two algorithms for estimating the mean lifetime are compared in a Monte Carlo experiment. In one method, the event time data is summarized as a histogram. The time endpoints of the histogram are selected so that the expected number of counts per bin contributed by the decay process, is constant. In the second method, the lifetime is estimated from the complete sequence of event times. The histogram method yields a less variable estimate of the mean lifetime. The optimal strategy for time allocation is found by minimizing the asymptotic variance of the lifetime (estimated from the pooled histogram data from all cycles) as a function T_{fill}, T_{decay} given knowledge of the filling rate of the trap and parameters which characterize the background process. The validity of the asymptotic approximation is demonstrated in Monte Carlo experiments.

In the histogram approach, estimates of the mean lifetime and signal and background paramaters are obtained by both a weighted least squares and a maximum likelihood method. For high count data, both methods yield estimates with similar properties. For low count data, the maximum likelihood approach yields estimates with lower bias.

A paper "Statistical planning for a neutron lifetime experiment using magnetically trapped

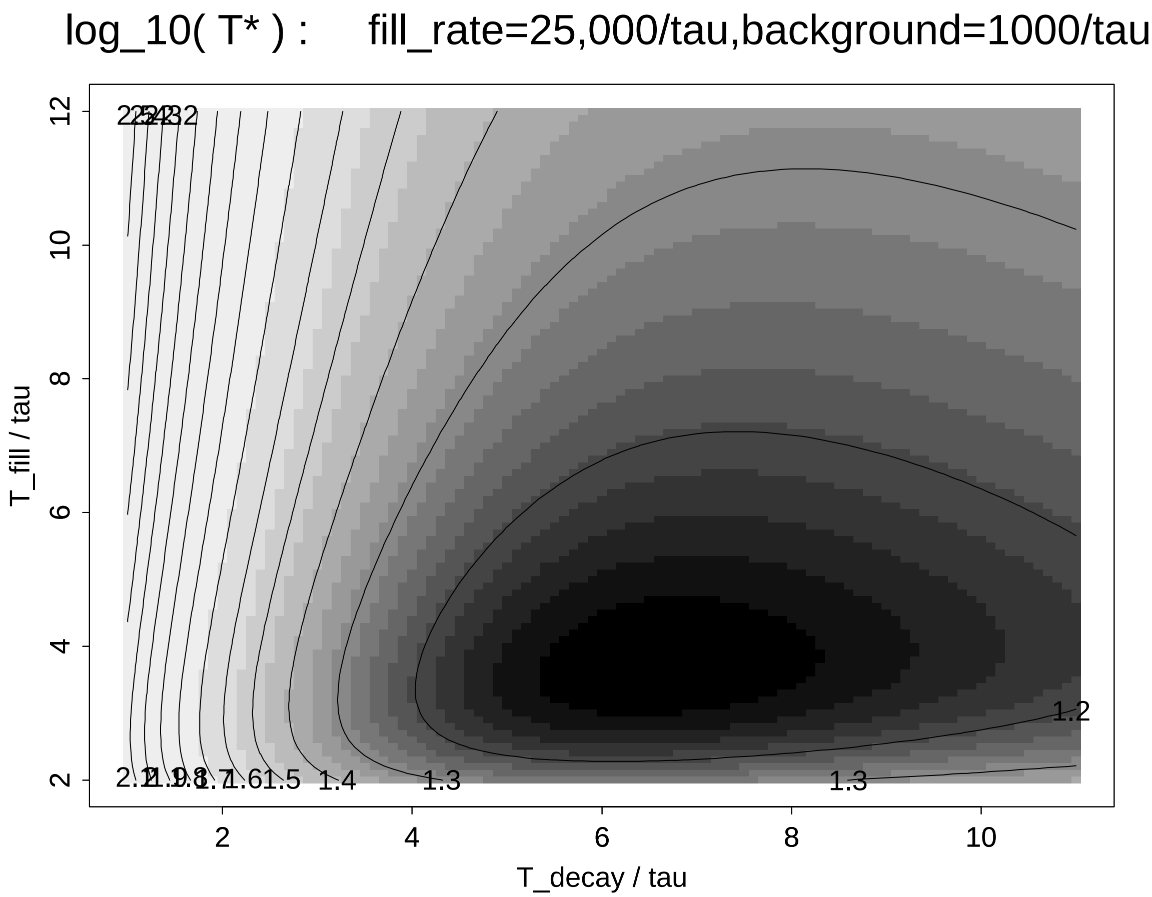


Figure 14: The approximate mean lifetime of the neutron is $\tau \approx 890\ s$. During the fill stage of each run cycle, the expected number of confined neutrons grows as $\lambda_{fill} * \tau(1 - exp(-T_{fill}/\tau))$ where λ_{fill} is the rate at which neutrons enter the trap, τ is the mean lifetime of the neutron and T_{fill} is the duration of the fill stage. We express the asymptotic standard error of the mean lifetime, estimated from data pooled from all run cycles, as $\frac{\sigma_{\hat{\tau}pool}}{\tau} \approx 0.001(\frac{T^*}{T_{total}})^{1/2}$ where the duration of the entire experiment is T_{total}. Above, $log_{10}(T^*)$ is plotted as a function of T_{fill}, T_{decay} for the case where $\lambda_{fill} = 25,000/\tau$ and the background is a stationary Poisson process with intensity rate equal to $1000/\tau$.

neutrons" will appear in Nuclear Instruments and Methods for Physics Research A.

3.3.2. Nonlinear Calibration of High Frequency Devices

Jack Wang, Kevin Coakley
Statistical Engineering Division, ITL

P. Hale and D. Larson
Optoelectronics Division, EEEL

We plan to calibrate photodiodes in the frequency domain up to 50 GHz. In the main experiment, a short laser pulse (50 fs) hits a photodiode. The output electrical signal is fed into a oscilloscope. The measured electrical signal is the convolution of the photodiode impulse response function and the oscilloscope impulse response function (plus noise). To characterize the photodiode, we need to know the impulse response function of the oscilloscope.

In a nose-to-nose oscilloscope calibration experiment, the kick out signal from oscilloscope i is fed into oscilloscope j. The measured pulse ($M_{ij}(t)$) is "approximately" the convolution of the oscilloscope impulse response function with itself. In the Fourier domain, the oscilloscope 1 impulse response function is

$$\hat{H}_1(w) = \sqrt{M_{12}(w)M_{13}(w)/M_{23}(w)}$$

. We divide the fourier transform of the measured electrical signal by the fourier transform of the oscilloscope impulse response function to get the transfer function of the photodiode in frequency space.

The nose-to-nose calibration data is affected by additive noise, random timing jitter noise and systematic timing errors known as time base distortion. Further, waveforms are affected by trime drift errors. Time base distortion errors can be estimated by complex demodulation. In this experiment, a sinusoid is fed into the oscilloscope. The output is demodulated. Alternatively, time base errors can be estimated using weighted least squares provided that multiple sinusoids at different frequencies are input into the oscilloscope. We also plan to measure time base distortion by time of flight methods with delta function pulses.

We plan to compare the nose-to-nose (magnitude and phase) characterization to alternative characterizations. One alternative characterization provides only 'magnitude" information. The other technique yields phase information using the Hilbert Transform. Finally, we plan to measure the absolute time delay (linear phase distortion) of the photodiode.

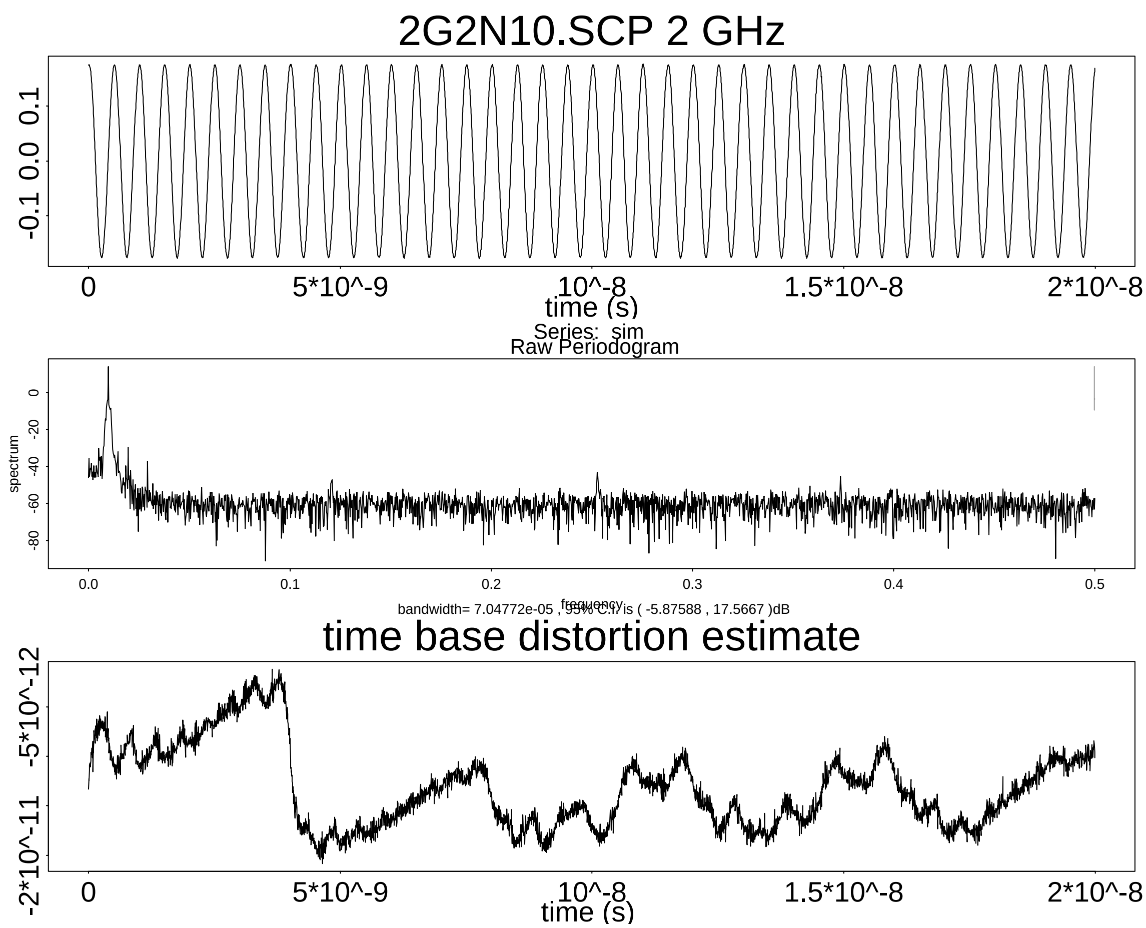


Figure 15: A 2 GHz sinusoid is measured by an equivalent time sampling approach (top). The actual and desired time of each sample is perturbed by a time base distortion error. We estimate the time base distortion by a complex demodulation algorithm in the frequency domain (bottom).

3.3.3. Detection and Quantification of Isotopic Ratio Inhomogeneity

Kevin J. Coakley
Statistical Engineering Division, ITL

David S. Simons
Surface and Microanalysis Science Division, CSTL

Most chemical elements in nature are multi-isotopic; i.e. they exist in several atomic forms with the same number of protons but different number of neutrons in their nuclei. Geologic and biological processes can alter the isotopic ratio of particular isotopes in a sample. Also, isotopic ratios can be intentionally altered by enrichment schemes. Materials with constant isotopic ratios are said to be isotopically homogeneous. In an inhomogeneous material, the isotopic ratio varies from location to location.

We quantify the spatial variation of the ratio of two isotopes within a material based on Secondary Ion Mass Spectrometry (SIMS) data. At many spatial locations, a detector counts each of two isotopes of a chemical element. At each location, we predict the less abundant isotope count in terms of the measured value of the more abundant isotope count and the estimated mean isotopic ratio. The difference between the measured and predicted value is divided by an estimate of its root mean square value. To estimate the spatial standard deviation of the isotopic ratio, we equate the sum of squared weighted residuals to its approximate expected value. The approximate expected value is obtained by a bootstrap resampling method. Based on the estimated null distribution of the estimated spatial standard deviation of the isotopic ratios, we test the hypothesis that the isotopic ratio is constant throughout the sample. To check the validity of our methods, we analyze SIMS data collected from a homogeneous Chromium sample. Results are consistent with the hypothesis of homogeneity. We simulate data corresponding to a sample where the isotopic ratio has a binary distribution. We find that when the standard deviation of the binary distribution exceeds twice the 86th percentile of the null distribution, detection of inhomogeneity is almost certain. Further, the estimated standard deviation closely tracks the actual standard deviation. To clarify results, the detection rate is expressed as a function of a scaled spatial standard deviation. The scaling factor is the sampling error associated with the esimated spatial mean value of the isotopic ratio. We submitted a paper to *Chemometrics and Intelligent Laboratory Systems*.

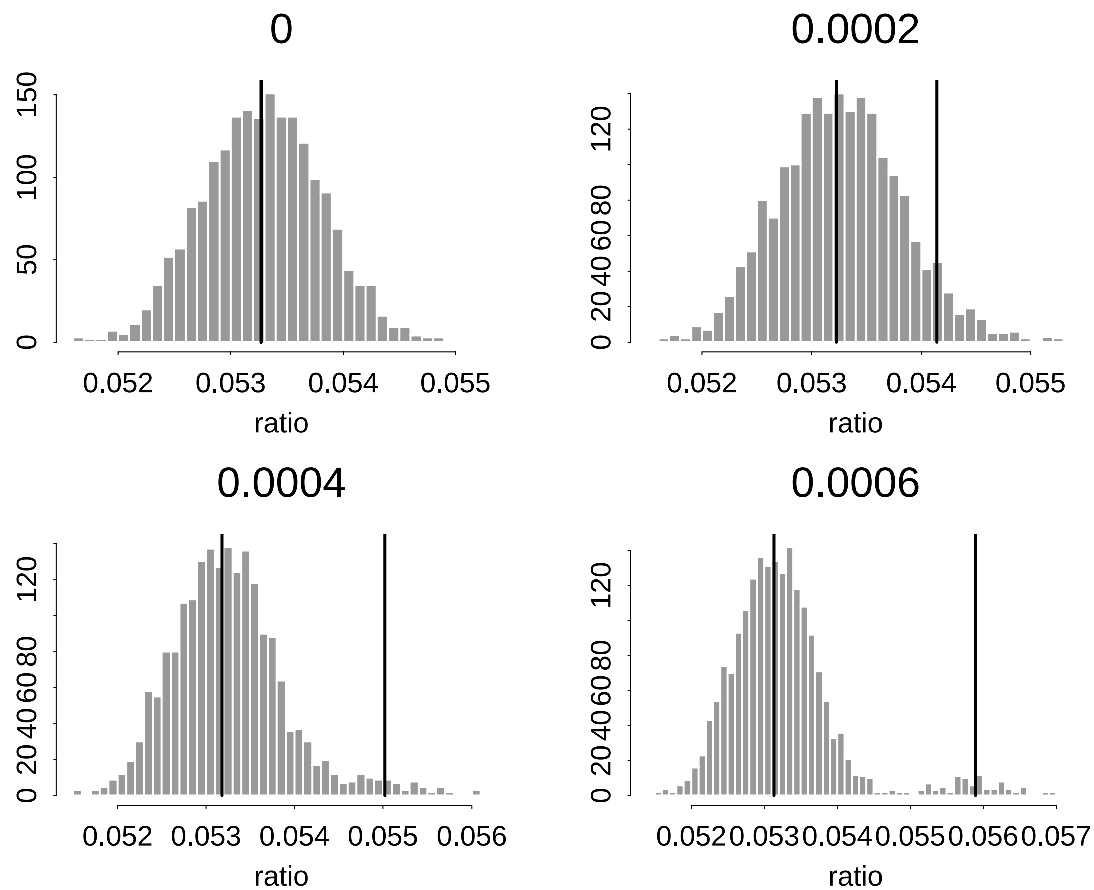


Figure 16: Sample histograms corresponding to simulated data where the isotopic ratio has a binary distribution. The standard deviation of the mixture distribution σ_r varies from 0 to 0.0006. The solid lines correspond to the values of the two isotopic ratios in the mixture. Mixing fractions are 0.95 and 0.05. For $\sigma_r > 0.0002$, for a test with size 0.10, the detection rate (of inhomogeneity) exceeds 99 percent and the estimated standard deviation $\hat{\sigma}_r$ closely tracks σ_r.

3.3.4. Consistency of Secondary Ion Mass Spectrometry and Neutron Depth Profiling Measurements

Kevin J. Coakley
Statistical Engineering Division, ITL

George Lamaze and Heather Chen-Mayer
Analytical Chemistry Division, CSTL

David S. Simons
Surface and Microanalysis Science Division, CSTL

Neutron Depth Profiling (NDP) is a nondestructive method for analysis of the concentration profile of an element in material. Inferences about the concentration depth profile are based on the observed energy spectrum of charged particles emitted due to specific nuclear reactions. The detector response function (DRF) is a probability transition matrix which relates the depth of emission to the expected energy spectrum of the detected particles. The DRF depends on the geometries of the emitter and detector, and assumed models for the stopping power of the material, energy straggling, multiple scattering and detector energy resolution.

In previous work, we developed a computer code to predict the DRF. Over the last year, we improved the model to account for non-Gaussian energy resolution functions. In a calibration experiment, the energy resolution function of the NDP detector was measured. For low energies, we model the energy resolution function as an exponential. For higher energies, we using a B-spline expansion.

A study of the consistency of NDP and SIMS data continued. To get more conclusive results, a new silicon sample was prepared. In the new sample, the boron profile is sharper than in previous samples. The depth profile of Boron in a Silicon sample was measured by Secondary Ion Mass Spectrometry (SIMS). Based on the measured SIMS profile and the modeled DRF, we predict the NDP energy spectrum.

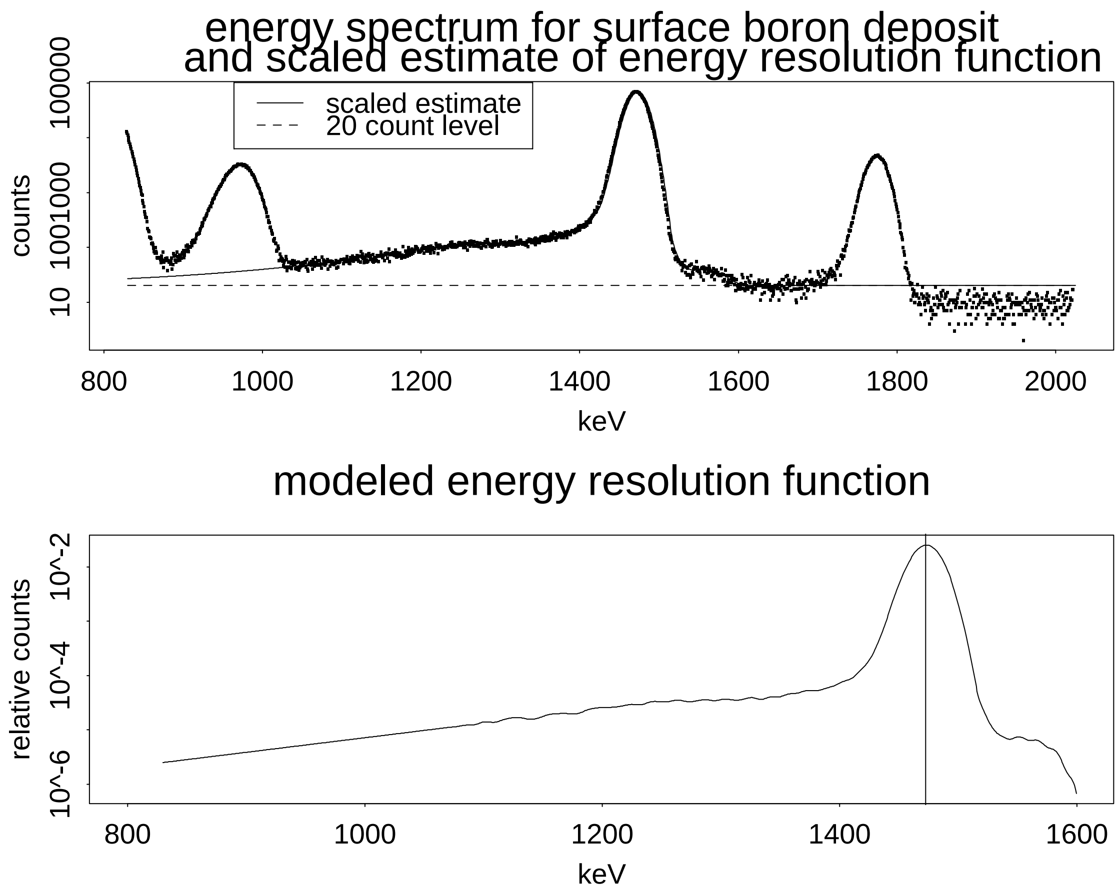


Figure 17: Based on the energy spectrum of alpha particles from a boron surface deposit (top), we esitimate the energy resolution function of the detector for a 1472.6 keV alpha particle (bottom). The vertical line is at 1472.6 keV. For low energies, we model the energy resolution function as an exponential. For higher energies, we used a B-spline approximation. In our modeling, we assume that the energy resolution function is shift invariant.

3.3.5. Modeling Constitutive Behavior of Steels

Kevin J. Coakley
Statistical Engineering Division, ITL

Yi-Wen Cheng
Materials Reliability Division, MSEL

Scientists in the Materials Reliability Division seek to improve the quality of sheet metal products manufactured by hot-strip rolling. The project is funded by the American Iron and Steel Institute and the Department of Energy. To achieve this goal, it is necessary to understand how a metal deforms under high stress. Based on experimental data collected at NIST, SED is developing a statistical model for predicting stress-strain behavior of metals as a function of chemistry, grain size, temperature and initial strain rate. The current model is an improved version of an earlier model developed at NIST.

The prediction variables in the new model are normalized so that the relative contribution of the different sources of variability are apparent. In the model, there are two terms in the prediction for stress. The first prediction term is a monotonically increasing function of strain. The second term represents a correction due to the dynamic recrystallization of the material. Due to this effect, stress is not necessarily a monotonically increasing function of strain.

Due to the high number of parameters in the model, the estimated parameters were obtained using a regularization approach. The model parameters are determined by minimizing a loss function which is the weighted sum of two terms. The first term is the sum of squared residuals. The second term is a penalty function. The model predicts the asymptotic value of stress for large values of strain. The penalty function is large when the predicted asymptotic value of stress is far from a prior estimate of the aysmptotic value. A weighting factor determines how much influence the penalty function has, relative to the sum of squared residuals term, in determining the parameter values. Estimates were obtained for various values of the weighting factor. Scientific judgement was used to select the best value of the weighting factor. Standard errors of the model paramters are estimated by bootstrap resampling.

The new model has better theoretical properties than does the earlier model. For certain choices of initial strain rate and temperature, the stress-strain curves should satisfy monotonicity constraints. For low to moderate strains, as strain is increased, the predicted stress curves for different grain sizes should not cross. However, the predicted stress curves from the older model did cross. In contrast, the predicted stress curves from the improved model do not cross.

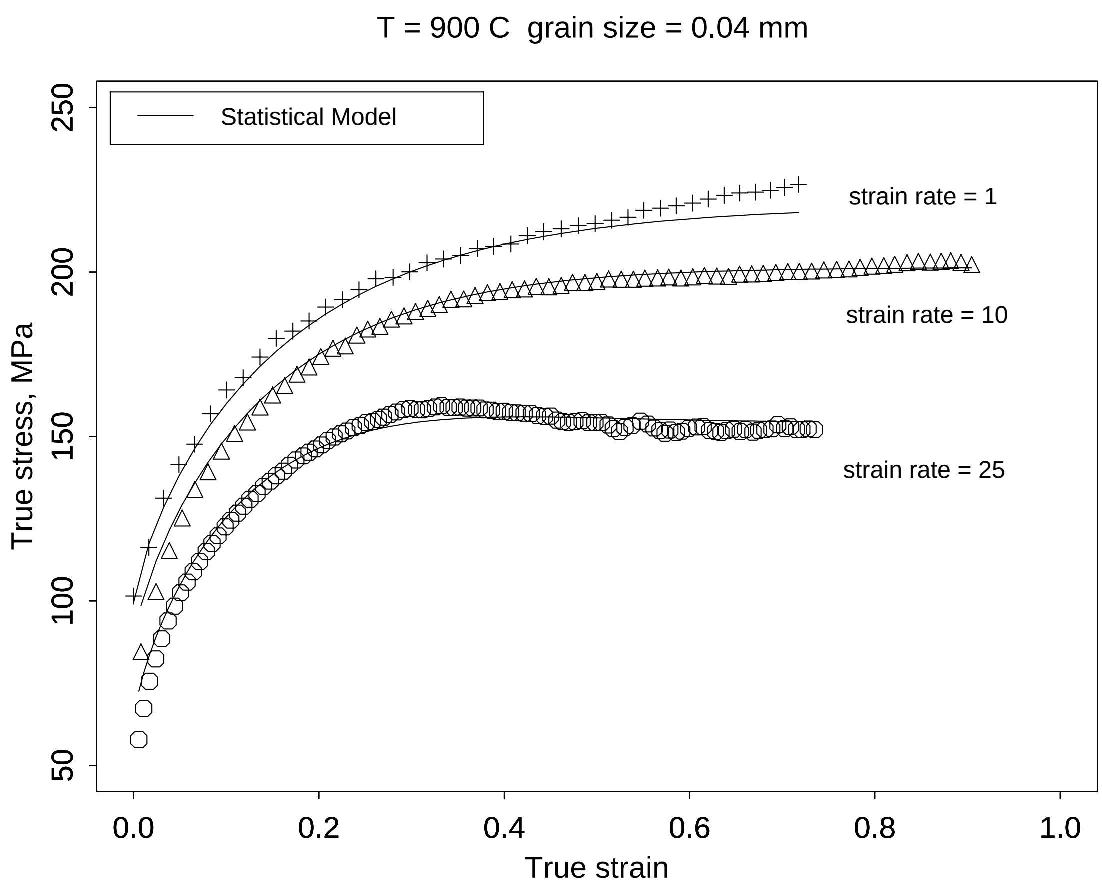


Figure 18: The measured stress-strain curves and a statistical model for the curves are compared for different experimental conditions. Due to dynamic recrystallization, strain is not a monotonic function of stress.

3.3.6. Nondestructive Determination of Residual Stress Using Electromagnetic-Acoustic Transducer

Kevin J. Coakley
Statistical Engineering Division, ITL

A.V. Clark
Materials Reliability Division, MSEL

Due to residual stress in a material, the velocity of sound depends on polarization state. Since the different polarization states travel at different velocities, interference occurs. As the propagation direction is varied relative to the polarization axes, the amplitude and phase of the transmitted wave varies. This is acoustic birefringence. Based on Electromagnetic-Acoustic transducer (EMAT) measurements of acoustic birefringence, stress can be determined.

In the experiment, the acoustic transducer is rotated. A sinusoidal signal enters the material and splits into two orthogonal polarization states. One state has a slow velocity of propagation. The other a fast velocity of propagation. In the ideal case, the measured signal is modeled as

$$s(t_z, z) \ = \ A\cos(k_s z - \omega t_z + \phi) \ + \ B\cos(k_f z - \omega t_z + \phi)$$

where $A \ = \ r\cos^2(\eta)$ and $B \ = \ r\sin^2(\eta)$. Above, η is the orientation of the slow velocity direction and the direction of the transducer. The transit time is t_z and the pathlength is z. Let

$$s \ = \ Re(w)$$

where w is complex. We have

$$w = A\exp(i\theta_1) + B\exp(i\theta_2)$$

where

$$\theta_1 = (k_s z - \omega t_z + \phi)$$

and

$$\theta_2 = (k_f z - \omega t_z + \phi)$$

The amplitude and phase of w are

$$|w| = \sqrt{A^2 + B^2 + 2AB\cos(\theta_1 - \theta_2)}$$

nd

$$PHASE(w) \ = \ ATAN(Im(w), Re(w)) \ = \ ATAN(A\sin\theta_1 + B\sin\theta_2, A\cos\theta_1 + B\cos\theta_2)$$

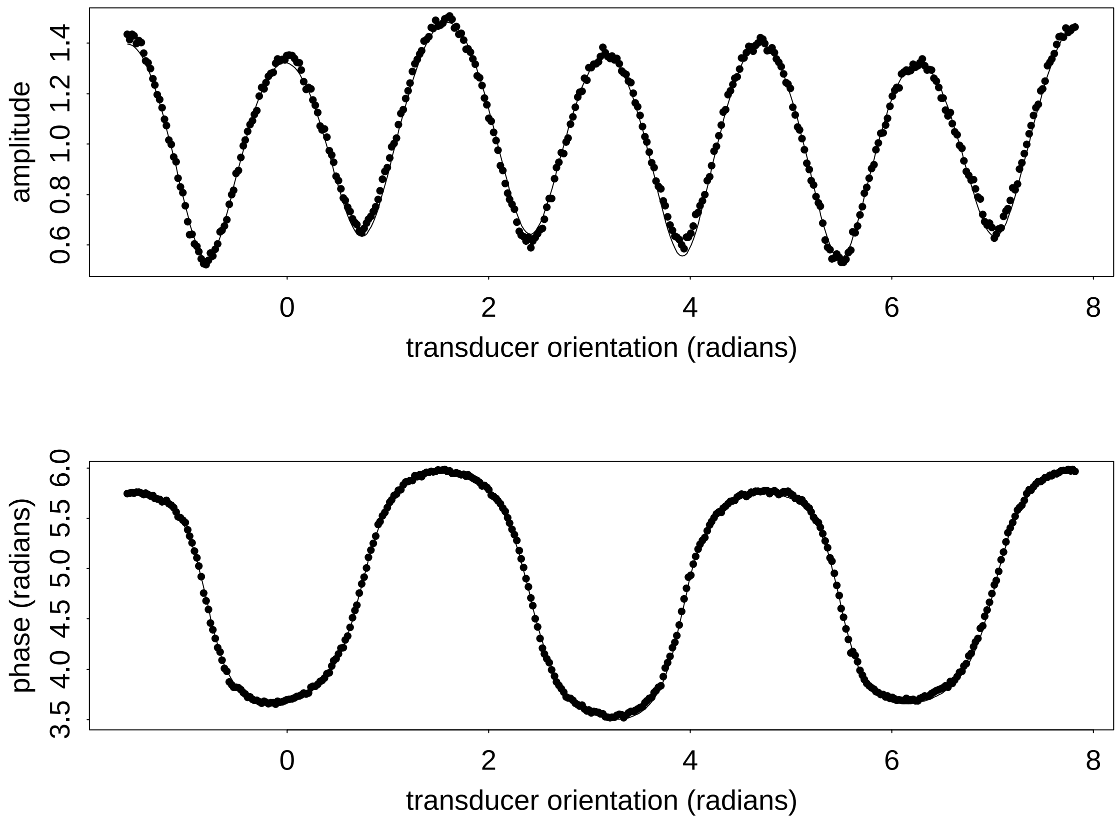


Figure 19: Amplitude and phase of acoustic echo is measured as the transducer rotates with respect to polarization axes. Predicted amplitude and phase is based on a model which accounts for material inhomogeneity and differential attenuation in steel.

We developed statistical models to estimate k_s and k_f from measured phase and amplitude. The model accounts for angle dependent effects due to material inhomogeneity and differential attenuation of the two polarization modes. A computer code for fitting the models to the data was developed for online data processing.

SED assisted in planning a study to compare EMAT and strain gauge measurements of stress. In the comparison study, ultrasound measurements will be made at many locations. However, strain gauge measurements will be made at just a few of these locations. We plan to compare ultrasound estimates with interpolated values of the strain gauge estimates. We developed a preliminary statistical model to predict the variance of the interpolated stress (from strain gauge measurements). Associated standard errors are also predicted.

3.3.7. Photo-Diode Interlaboratory Comparison

James J. Filliben
Statistical Engineering Division, ITL

Steve Brown
Tom Larason
Sally Bruce
Optical Technology Division, PL

As part of its leadership role in the international optical physics community, the Optical Technology Division (844) of the Physics Laboratory will be conducting an extensive interlab experiment to determine the responsivity (power output) and homogeneity of photo-diodes (= photo detectors).

The photo-diode measurements are non-destructive and so diodes can be reused, rechecked, and recirculated. A serious complicating issues is that the diodes themselves may drift.

This experiment design contends with the usual mix of experimental objectives and practical constraints. The experimental objective is to determine the equivalency of participating labs for the above two responses. Practical constraints include: 1) How many diodes NIST could afford to buy? 2) How many diodes NIST could afford to run as a control? 3) How long the experiment was to last in toto?

Statistical issues involve: 1) How many diodes to send to each lab? 2) How many replicates should a lab run? 3) Exactly what diodes a lab should receive? 4) How many recyclings (the diodes get returned to NIST, remeasured, and redistributed)? 5) Should diodes be returned to the same lab or sent to other labs? 6) How to assess within-lab variability problems? 7) How to assess within-lab drift problems? 8) How to assess between-lab bias problems? 9) How to detect and correct for diode drift?

A series of admissible designs were constructed and evaluated. A metric was formulated to evaluate/score a given design: 1) the number of diodes that a given lab sees; 2) the number of labs that a given diode "sees"; 3) the number of replications of a given diode within a lab; and 4) the total number of physical diodes.

The theoretical "ideal design" will have the property that 1) metric 1 is large (to minimize confounding); 2) metric 2 is large (to maximize between-lab discrimination); 3) metric 3 is large (to maximize within-lab discrimination); and 4) metric 4 is relatively small (to be affordable).

An unconstrained optimzation of all 4 metrics is not possible from a practical point of view–tradeoffs exist. A schematic of the final design is presented on the opposing page.

Photo-diode Interlab Experiment

Lab \ Pass	1	2	3	4
1 2 3 4 5	Diode 01 02 : 03 04 05 : 06 07 08 : 09 10 11 : 12 13 14 : 15 A : A'		B A'	
6 7 8 9 10	16 17 : 18 19 20 : 21 22 23 : 24 25 26 : 27 28 29 : 30 B : B'		A B'	
11 12 13 14 15		A A'		B A'
16 17 18 19 20		B B'		A B'

NIST	Control Run	Control Run	Control Run	Control Run	Control Run

Constraints: diodes <= 30 labs <= 20 runs <= 120
control points <= 5

Characteristics: diodes/lab = 6 labs/diode = 4

Figure 20: Experiment design for photo-diode interlab experiment.

3.3.8. Semiconductor Bond Strength

James J. Filliben
Statistical Engineering Division, ITL

Corrine Mansfield
Green Tweed Corporation

Mansfield attended SED's experiment design for industry 5-day course: Improving Product and Process Quality Using Experiment Design given by Eric Lagergren, Lisa Gill, and James Filliben on December 1 5, 1997 at NIST. Green-Tweed Corporation is a semiconductor-related company in Pennsylvania. The goal of Mansfield's project was to develop stronger bond strengths in certain semiconductor components. Mansfield had 8 factors under investigation: 1. cleaning agent; 2. surface roughness; 3. pre-heat time; 4. blast-to-glue time; 5. glue dilution; 6. glue-to-mold time; 7. molding condition; and 8. protection coat.

Based on the course, and the in-class consultation that was part of the course, Corinne had constructed an 8-factor, 16-run, 2-level fractional factorial design which is an excellent, efficient design for her problem. Because of certain complexities in the 8-factor design that was actually run (she chose the non-Yates order design on page 402 of Box, Hunter, & Hunter: Statistics for Experimenters), Corinne was a bit unsure as to the validity and completeness of the conclusions from her analysis.

After making adjustments for the non-Yates confounding structure of her design, we ran a routine re-analysis of her experiment. Her conclusions were reaffirmed in some cases and augmented in others. With respect to the desired increase in bond strength, the important factors were ferreted out, best settings which will yield consistently higher bond strengths were identified, and an empirical model involving the dominant factors and interactions was constructed.

The attached graphic illustates one of the more important analysis techniques for 2-level factorial designs: an interactions effect matrix. To protect the proprietary nature of this industrial experiment, we have scrambled the 8 factors and coded them on the plot.

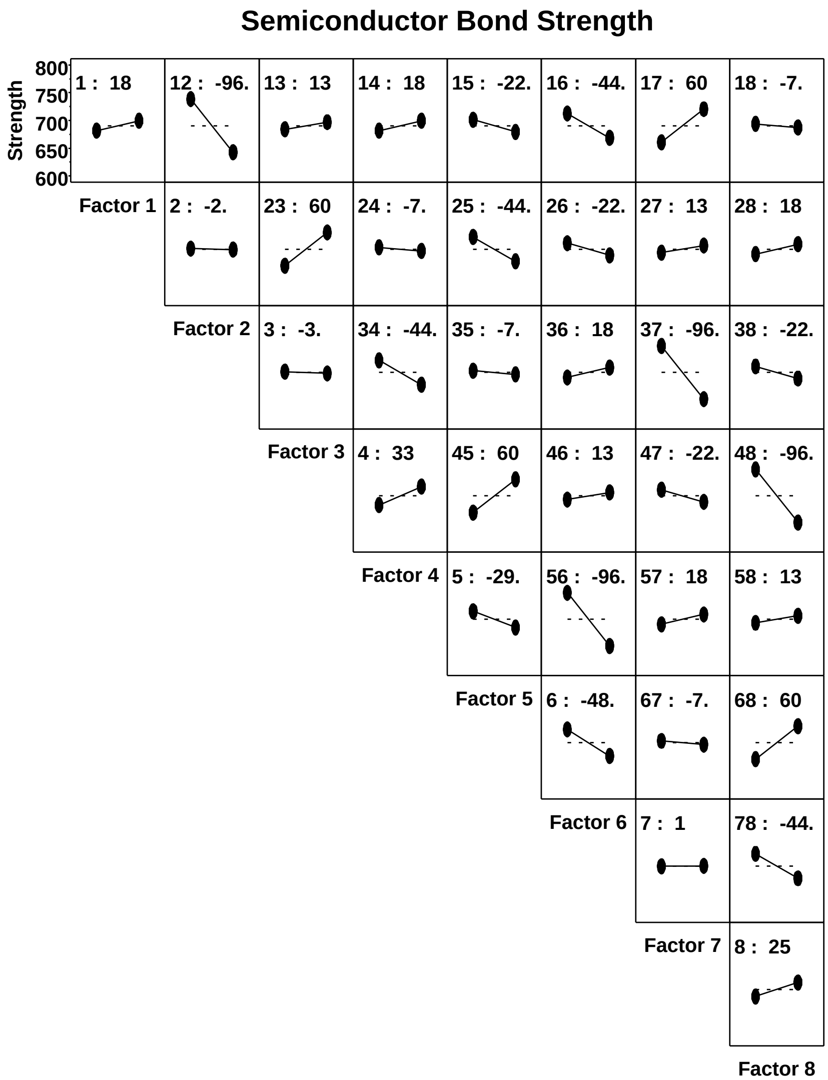


Figure 21: This interaction effects plot matrix shows main effects along the diagonal and 2-term interactions off-diagonal. Among the main effects, note the relative importance of factor 6. Among the 2-term interactions, note the intrinsic confounding structure for this 2**(8-4) as revealed by equivalent off-diagonal plots (e.g., the 1-2, 3-7, 5-6, and 4-8 interactions, and the probable reality of the 5-6 interaction.

3.3.9. Performance Evaluation for Lead-in-Paint Measuring Devices Under Simulated Field Conditions

Eric S. Lagergren
Statistical Engineering Division, ITL

Susannah B. Schiller
Advanced Technology Program

Mary E. McKnight
Building Materials Division, BFRL

In January 1995, USA Today reported: "Federal Housing officials have ordered retests to detect toxic lead paint in 85 public housing projects where hundreds of millions of dollars in tests may have been flawed. At issue is whether tests by portable X- ray machines are reliable." As part of an effort to improve the reliability of lead-in-paint measuring devices, the U.S. Dept. of HUD asked NIST to identify and quantify factors affecting the field performance of these portable X-ray fluorescent (XRF) devices. The ultimate objective of this study is to develop a protocol for assessing field precision and bias of XRF instruments that measure lead concentration in painted surfaces.

The protocol would consist of taking XRF measurements on lead-in- paint standards at a specified set of noise conditions known to cause variability in XRF field measurements. The noise conditions are combinations of settings of "noise" factors known to cause variability in field measurements. A candidate list of noise factors can be generated. However, the order of importance of these candidate noise factors, i.e., which cause greatest measurement variability, is currently unknown. A laboratory experiment was conducted to identify the most important noise factors. In this experiment, the noise factors were systematically varied according to a statistical experimental design to study their effect on XRF measurements. Of course, the conclusions from this lab experiment must be validated in the field to ensure that all important sources of variability have been captured. In this experiment, eight noise factors were studied at each of two settings using a 16-run (out of a possible $128 = 2^8$) fractional factorial design. This well-chosen subset of 16 runs permits free and clear estimation of the primary effects of all factors and limited information on two-way interactions between factors. The 16 noise factor conditions were studied for each of four XRF instruments and two lead concentrations.

The figure shows that x6, the distance of the instrument from the surface, is the dominant noise factor, followed by x3, the underlying substrate (wood or steel). Substrate has a large effect for only two of the instruments, indicating that the other two invoke a substrate correction. The "distance-from-surface" noise factor was included to simulate non-flat surfaces such as wood molding, metal pipe, and stucco. A follow-up experiment confirmed that the "distance-from-surface" factor is indeed an adequate surrogate for non-flat surfaces.

A sample test protocol was developed in the important noise factors and used to assess the field uncertainty of XRF measurements. A report for HUD is currently underway.

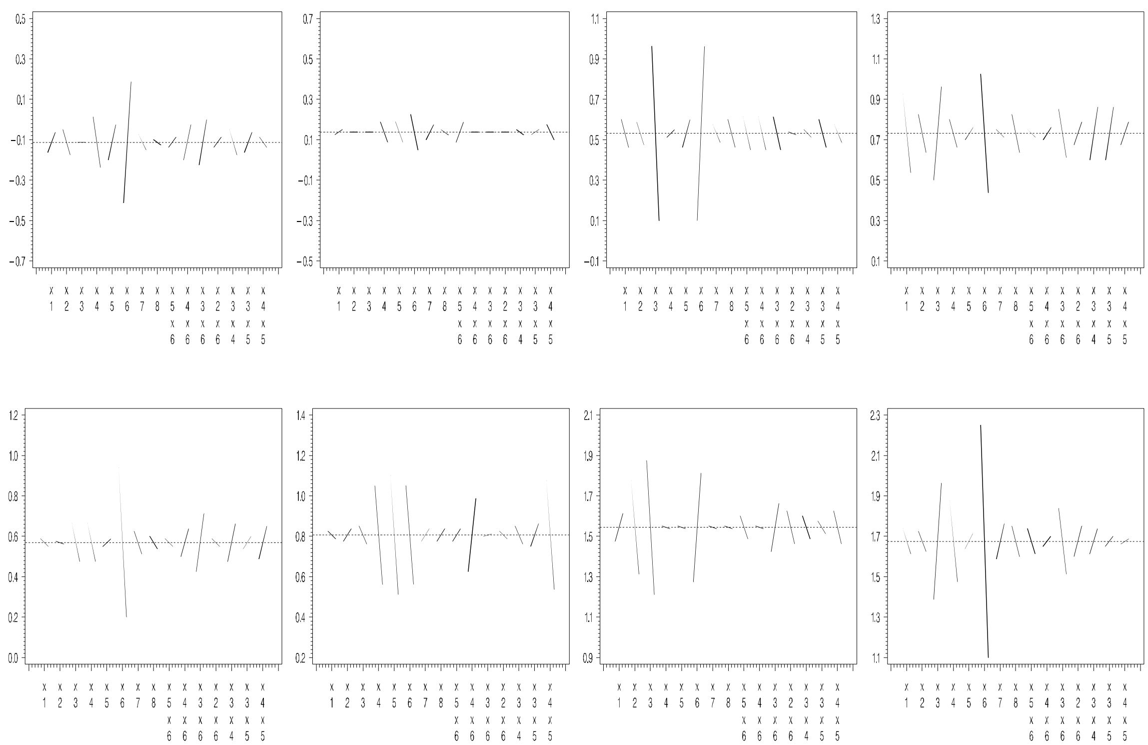


Figure 22: Plots of mean XRF response versus factors for instruments from four manufacturers (columns) and two lead levels (rows).

3.3.10. Optimizing High-Performance Concretes Using Mixture Designs

Eric S. Lagergren
James J. Filliben
Lynne B. Hare
Statistical Engineering Division, ITL

Dale P. Bentz
Ken A. Snyder
Building Materials Division, BFRL

Marcia Simon
Federal Highway Administration (FHWA)

Optimizing high-performance concrete is currently more of an art than a science. Some guidelines are available for selecting optimal conditions, but no systematic approach is used to identify these conditions. As a result, trial and error or "one-factor-at-a-time" designs are typically used to identify best mixtures. A collaboration is underway between the FHWA and NIST's Statistical Engineering and Building Materials Divisions to investigate the feasibility of using mixture design and analysis techniques for optimizing high-performance concrete. A second objective is to develop a World Wide Web service for users to design and analyze mixture experiments for optimizing concrete mixes.

In the first phase of this work, a laboratory experiment was conducted to study six mixture components: water, cement, fine and coarse aggregate, superplasticizer, and microsilica. The first four components produce concrete. The last two enhance specific properties yielding "high-performance" concrete. The properties of interest are workability, air content, strength, and chloride ion permeability. Since the proportions of the six components were constrained to a subset of the full mixture space, standard Scheff simplex designs could not be used. Instead, a modified distance-based design was used. This design permits fitting a second-order Scheff polynomial and includes points for checking the adequacy of the fitted model and estimate repeatability.

The figure shows the contour plot of the fitted model for 28-day strength over the simplex in three components: superplasticizer (HRWRA), coarse aggregate and fine aggregate at fixed proportions of water, cement, and microsilica. The plot shows that increasing the proportion of fine aggregate (moving toward the vertex labelled fine aggregate) increases 28-day strength. Optimal conditions were identified for each property using the contour plots and numerical optimization routines. Also, a best condition across multiple responses was found using desirability functions.

There is some concern that the mixture design & analysis approach is more complicated than necessary for novice users, the customers of the proposed web service. We are currently investigating other approaches using more balanced response surface designs that would lend themselves to simpler, graphical analyses and be less reliant on model fitting.

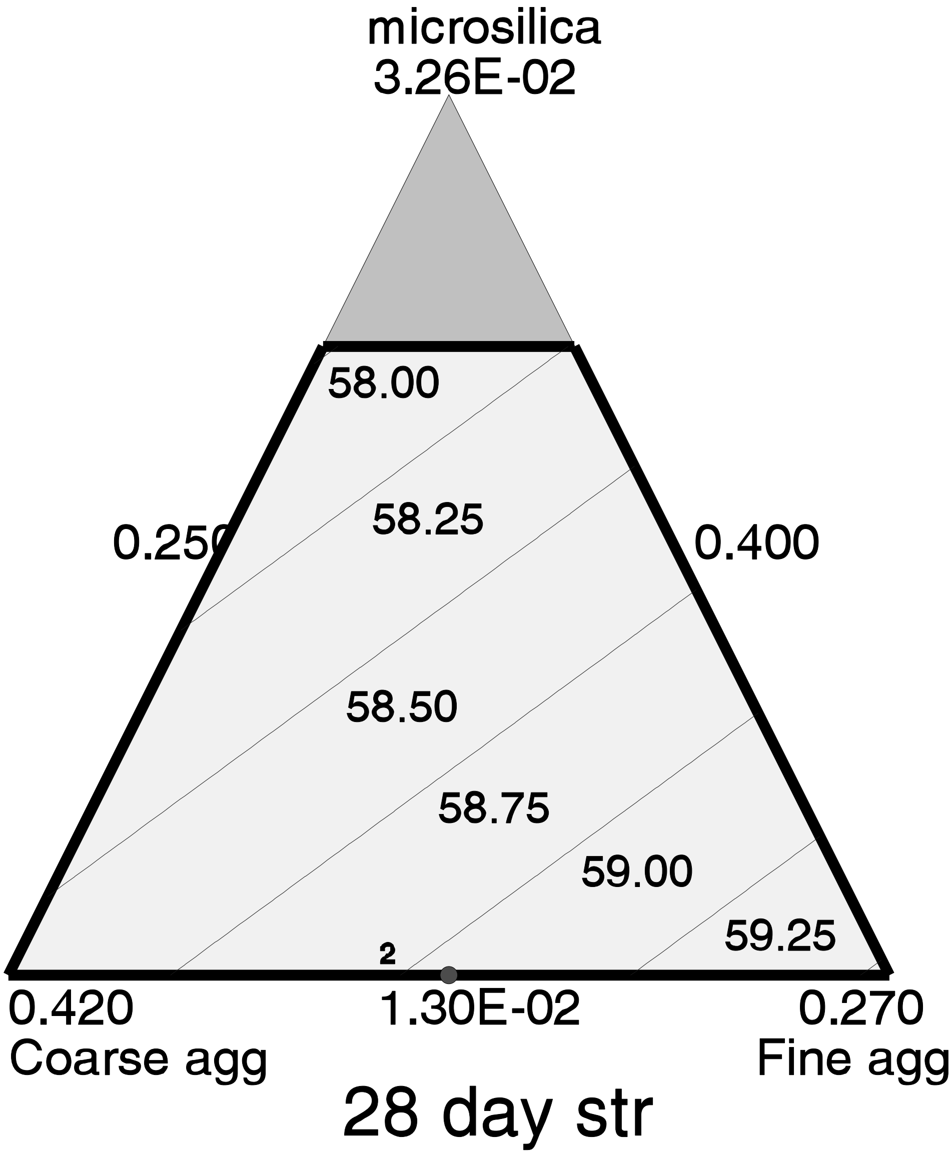


Figure 23: Contour plot for 28-day strength (MPa) in microsilica, coarse aggregate, and fine aggregate at fixed proportions of water (0.16), cement (0.15), and superplasticizer (0.0074).

3.3.11. Creep-Rupture Performance of Adhesively-Bonded Roofing Seams

Mark G. Vangel
Statistical Engineering Division, ITL

Walter J. Rossiter
Building Materials Division, BFRL

Adhesively-bonded EPDM (a rubber material) is widely used for low-slope industrial roofing. There are two main types of adhesive systems for seams on these roofs: a liquid adhesive, and several varieties of tape adhesive. Liquid adhesive is widely used, but it is volatile and relatively expensive to apply. An objective demonstration that tape adhesives are at least as reliable as the liquid will greatly increase the use of these adhesives. A consortium of NIST, professional roofing trade associations, and roofing adhesive manufacturers was formed, in part, to perform such a study.

The chosen measures of performance for the experimental seams are strength and creep lifetime, i.e. the time-to-failure under a constant load. In Phase 1 of this investigation, specimens from two tape systems and a liquid adhesive were tested in creep-rupture at various loads. The main conclusion of this phase was that the tape-bonded seams appear to perform at least as well (in terms of creep life) as adhesive-bonded seams. In Phase 2, the influence of application factors on performance was investigated.

Phase 3 of this study consists of five experiments, investigating the following:

1. Effects of elevated temperature during loading,
2. Effects of elevated temperature exposure during fabrication,
3. Effects of exposure to industry-developed protocols,
4. Effect of cold temperature during fabrication, and
5. Seam response in shear.

The responses measured included strength and creep life.

The figure displays the strength data from the second experiment, during which roof seam specimens were prepared at various elevated temperatures, and left at these temperatures for varying times. There are several interesting patterns in these data, the most obvious of which is that both tape adhesive systems appear to be stronger than the liquid adhesive system.

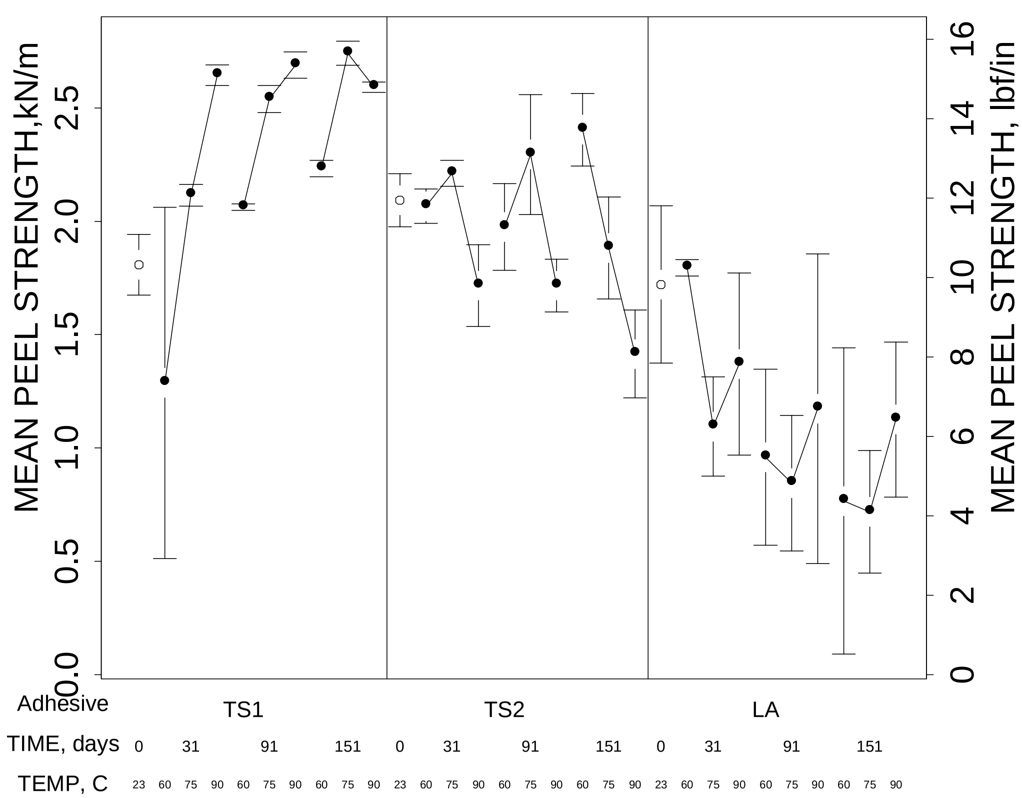


Figure 24: Strength as a function of elevated temperature and time at temperature of tape (**TS1**, **TS2**) and liquid (**LA**) adhesively-bonded EPDM seams. Open circles identify room-temperature control tests.

3.3.12. Volume Recovery of Poly(Vinyl Acetate)

Mark G. Vangel, Andrew L. Rukhin, Stefan D. Leigh
Statistical Engineering Division, ITL

Gregory B. McKenna
Polymers Division, MSEL

B. Lotz, C. Straupe *Institute Charles Sadron, Strasborg, France*

When polymer glasses are equillibrated at an initial temperature (T_i) and then placed in a water bath at a second temperature (T_f), the volume of the polymer will change smoothly until the material attains the new equilibrium temperature. A measure of how 'far' the volume ($v(t)$) of a specimen is from equilibrium (v_f) at time t is $\delta(t) = (v(t) - v_f)/v_f$. The negative of the derivative of the logarithm of $|\delta(t)|$, a measure of the rate of approach to equilibrium, was defined as the 'effective' recovery time, τ_{eff}, by A.J. Kovacs in a very influential 1964 article summarizing an extensive experimental program on volume recovery of poly(vinyl acetate).

Kovacs claimed that when approaching the same equibrium temperature T_f from different initial temperatures, say T_i and T_i', the τ_{eff} values differed for δ values as small as could be reliably measured. This is somewhat paradoxical, since it suggests that the specimen 'remembers' T_i when it is close to a very different T_f. A recent publication by a prominent researcher has called this assertion into question by arguing that Kovacs' experimental uncertainty is much larger than he realized.

It is of considerable theoretical importance to establish to what extent Kovacs was correct, and we have addressed this question by means of a thorough statistical analysis of 96 experiments done by Kovacs and his students, for poly(vinyl acetate) at many initial and final temperatures. We conclude that Kovacs was essentially correct in his assessment of experimental uncertainty, and that the recent article critical of Kovacs experimental work overstates this uncertainty, primarily by ignoring the positive correlation among measurements of $\delta(t)$ made close together in time.

We used a propagation-of-errors argument to express the uncertainty in divided-difference estimates of $\tau_{\text{eff}}(t)$ in terms of the uncertainty in measurements of $\delta(t)$, and the correlation among these measurements. The uncertainty in $\delta(t)$ was bounded above using physical considerations, and the correlation among measurements of $\delta(t)$ was estimated from Kovacs' many replicated experiments. Together, these results enabled us to calculate approximate confidence intervals on $\tau_{\text{eff}}(t)$, for curves having different T_is, but the same T_f. From these confidence bands one can quantify how small $\delta(t)$ needs to be in order to conclude that the τ_{eff} values are not statistically distinguishable. This analysis is illustrated in the figure for a pair of experimental curves.

These quantitative results were corroborated by a second statistical analysis, using a repeated measures model, and by a qualitative graphical analysis.

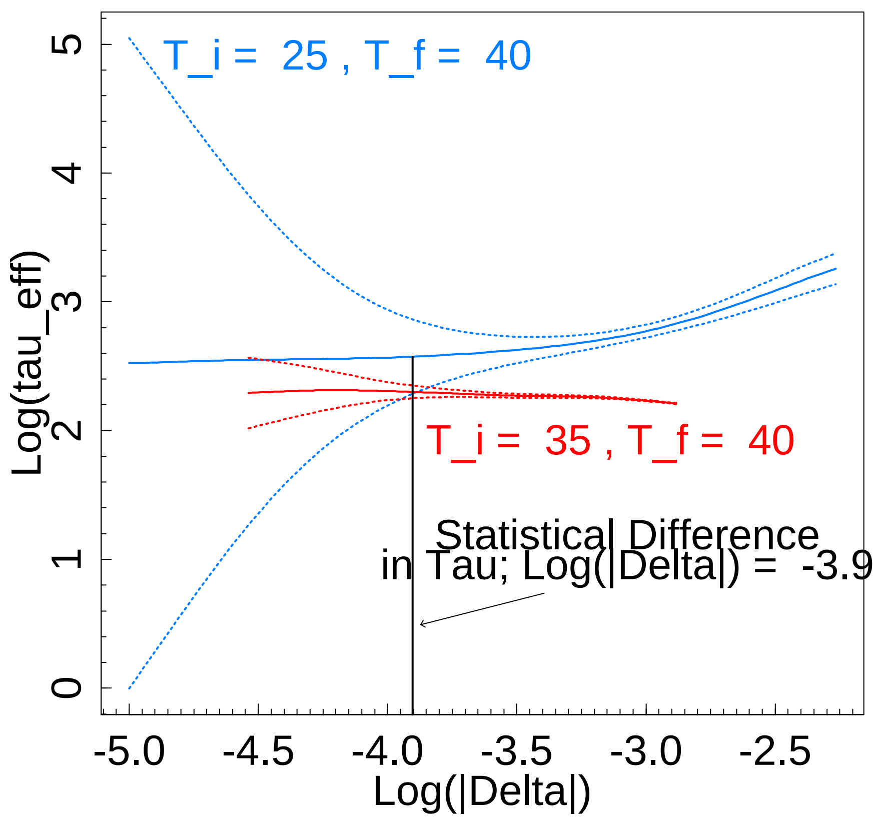


Figure 25: Estimated $\log(\tau_{\text{eff}})$ with 95% individual confidence bands, as functions of t, for initial temperatures $T_i = 25°\text{C}$ and $T_i' = 35°\text{C}$, and final temperature $T_f = 40°\text{C}$. The curves were estimated from A.J. Kovacs' experimental data on volume recovery of poly(vinyl acetate).

3.4. Statistical Inference

3.4.1. Semiconductor Growth Rate

James J. Filliben
Mark G. Vangel
Statistical Engineering Division, ITL

Mike Tietjen
Semiconductor Electronics Division, EEEL

In the semiconductor industry there are 3 common ways of estimating semiconductor growth rate (which in turn allows accurate achievement of pre-specified wafer thickness and accurate estimation of wafer proportional composition):

1) spectroscopic ellipsometry (which makes use of light scattering); 2) x-ray diffration (which makes use of x-rays); 3) RHEED (= Reflective High Energy Electronic Diffraction) (which makes use of electron gun diffraction).

Of the three, the RHEED method is most heavily used; lack of an accuracy estimate for the growth rate from RHEED is an industry impediment to the precise growing of wafers.

Mike Tietjen (and Joe Pelligrino) of the Semiconductor Devices Division of EEEL are addressing the problem of RHEED method accuracy. In particular, they are assessing and determining the accuracy of the RHEED method for measuring growth rate and composition of 3-element (aluminum-gallium-arsenic) III-IV semiconductor wafers.

Project questions include: 1) How accurate is that growth rate estimate? 2) How to set up a designed experiment to arrive at a robust estimate of growth rate? 3) How to convert the growth rate into estimates for the composition percentages (there are 2 complementary formulas)? 4) How good are the composition estimates? 5) How many replicates are needed?

SED contributed to this project in the following fashion: 1) An appropriate designed experiment was constructed; 2) The virtues of a spectral analysis (as opposed to a Fourier analysis) were passed on. 3) Growth rate estimates based on the spectrum and on FFT were carried out. 4) The industry software FFT was failing under certain (low-frequency) circumstances. SED/Dataplot provided such estimates. 5) Mike had some concern about the accuracy of the industry-software FFT algorithm. Numerically accurate estimates of the FFT were provided (in the spirit of StRD). 6) Precision values were computed for the growth rate. 7) Composition percentage estimates were computed. 8) Propagation of error was carried out to arrive at uncertainty estimates for the composition percentage estimates.

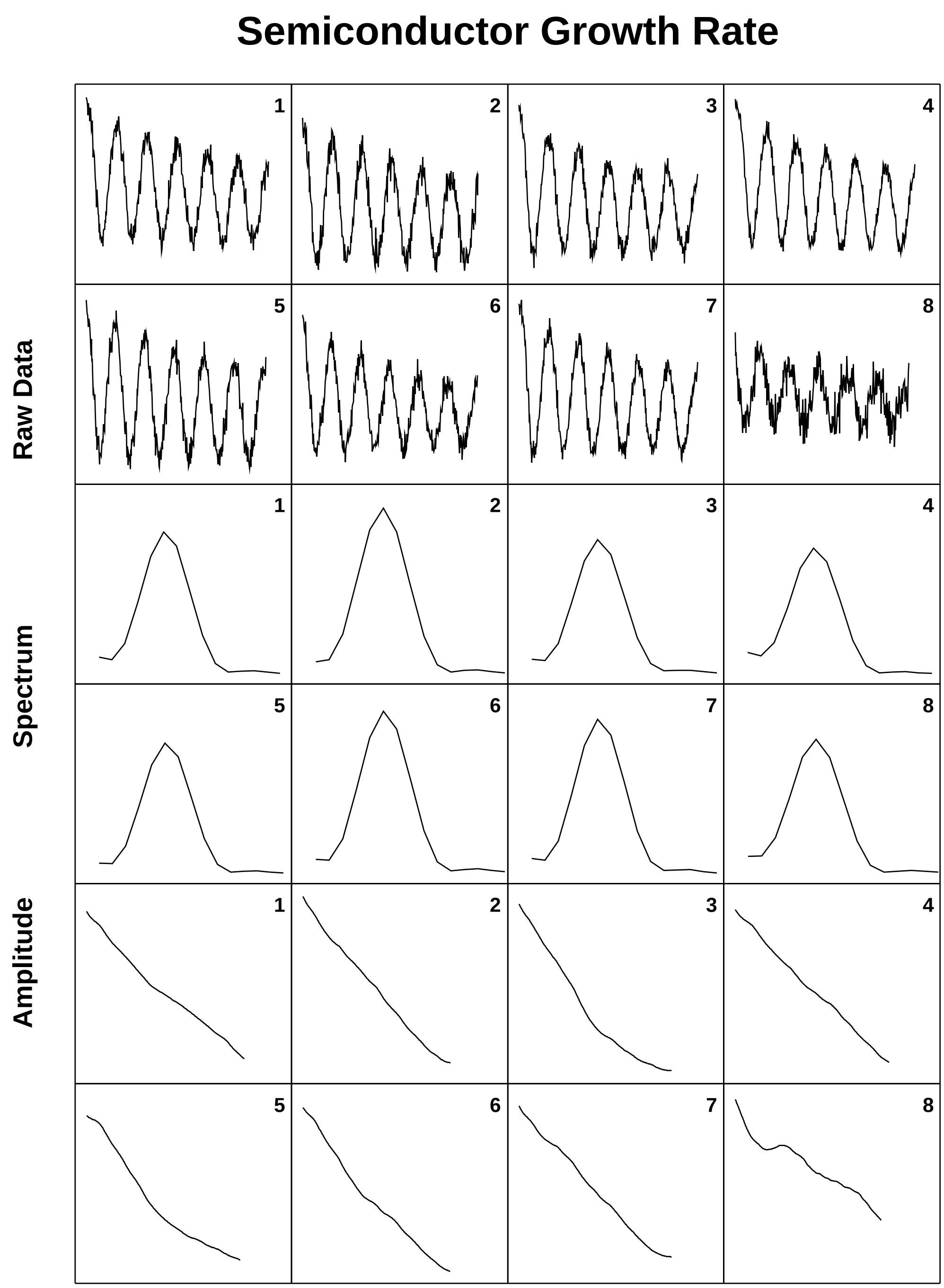


Figure 26: The first set of plots (raw data traces) shows the characteristic sinosoidal nature of the deposition process. The second set of plots (spectral plots) shows the consistency of the dominant frequency. The third set of plots (complex demodulation amplitude plots with demodulation frequency .016 cycles per observation) shows the linear damping of the process.

3.4.2. Electromigration & Reliability in Semiconductor Thin Film Interconnect Stripes

James J. Filliben
Alan Heckert
Statistical Engineering Division, ITL

Santos Mayo
Semiconductor Electronics Division, EEEL

Semiconductor manufacturers have made great strides in increasing chip speed while maintaining chip reliability. An increasingly important potential size and speed bottleneck is not the chip, but the aluminum alloy thin film stripes that interconnect one chip with another. For the system to be speed-efficient as a whole, it is important that the interconnects–not just the chips–be made smaller.

Such thin film interconnect miniaturization suffers from electromigration: the phenomenon in which ion-electron collisions intrinsic to current flow induce internal physical movement causing physical deformations and voids being formed and de-formed. Such migration has of course a deleterious effect on the lifetime and reliability of the interconnect. The understanding of such microstructure mechanics is essential for future reliable miniaturization.

Santos Mayo of the Semiconductor Electronics Division of EEEL is addressing this problem. Specifically, he is investigating and characterizing the damage/reliability mechanisms associated with electromagnetic processes in aluminum alloy thin film interconnect stripes on a silicon dioxide substrate.

Santos has carried out an accelerated experiment whereby he has applied high temperature (in the 175 to 200 degree C range) and high current (in the 1.5 to 2 million amperes per square centimeter range) to such thin film interconnect stripes over the course of several hours and recorded the resistance trace at .001 second intervals. Such a resistance trace is relatively uneventful for most of the time period. At "random" points in time, however, the traces have resistance surges which are signatures of certain types of electromigration activities (void formation, void dispersion, etc.) and which cumulatively serve to degrade the current-carrying capacity of the stripe and ultimately shorten the lifetime of the stripe. Santos' data collection system has been designed to ignore periods of no activity (when a Santos-defined threshold has not been exceeded, but to heavily record (1000 observations per second) when a "possible-event" has been detected. Such several-hour data sets typically contain about 200,000 observations–200 "possible-events" with 1000 observations per "possible-event".

SED is analyzing such records for event/peak identification; peak area calculation; and categorization of events into different types (corresponding to different electromigration mechanisms). A future experiment design will check the generality of the conclusions.

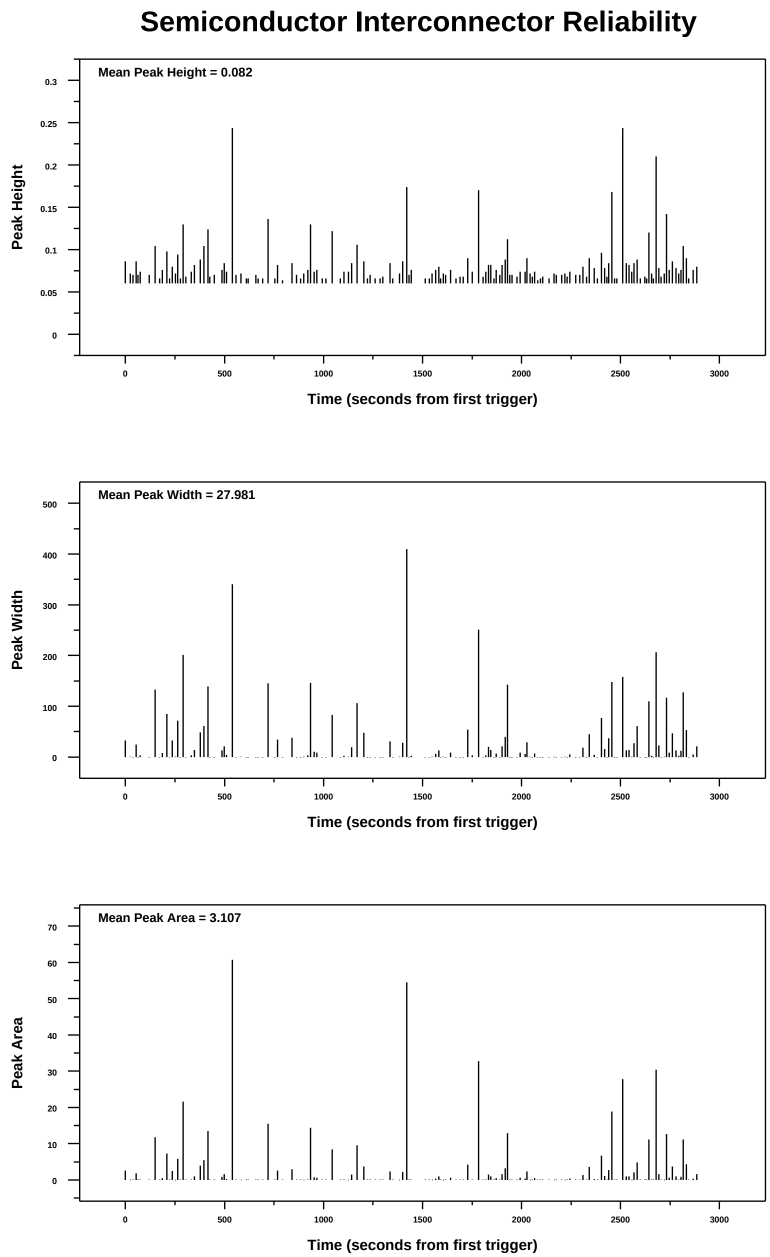


Figure 27: Height, width, and area of event peaks.

3.4.3. High-dimensional Empirical Linear Prediction

Hung-kung Liu
William F. Guthrie
Statistical Engineering Division, ITL

J.T. Gene Hwang
Cornell University

Michael T. Souders
Gerard N. Stenbakken
Electricity Division, EEEL

Many engineering problems involve high-dimensional observations with mean vectors sitting in a lower dimensional space. Exhaustive measurement of all the elements of an observation is often time consuming and expensive. Applying a traditional multivariate linear model, one can incorporate a small number of the elements of the observation with a known design matrix to predict the rest of the elements. However, for a complicated engineering system, the design matrix is often hard to be fully determined. We investigate an empirical linear model, in which we allow ourselves to use the data to determine the size of the design matrix and to estimate the unknown part of the design matrix. This estimated model is then used to construct point and interval estimates for the future observation. This technique is called HELP (High-dimensional Empirical Linear Prediction).

As an example, for a 13 bit A/D converter, to be absolutely sureof performance, one needs to test 8192 outputs, corresponding to transition levels (usually voltage levels) for the conversion of the analog signals to the digital signals. By using the exhaustive measurements of 88 converters and the measurement of only 64 transition levels of a future converter, HELP predicts well the behavior of the rest 8128 transition levels. In manufacturing converters, testing cost constitutes 20% to 50% of the total manufacturing cost. Obviously, a reduction from 8192 measurements down to 64 measurements, less than one percent, can reduce production cost tremendously.

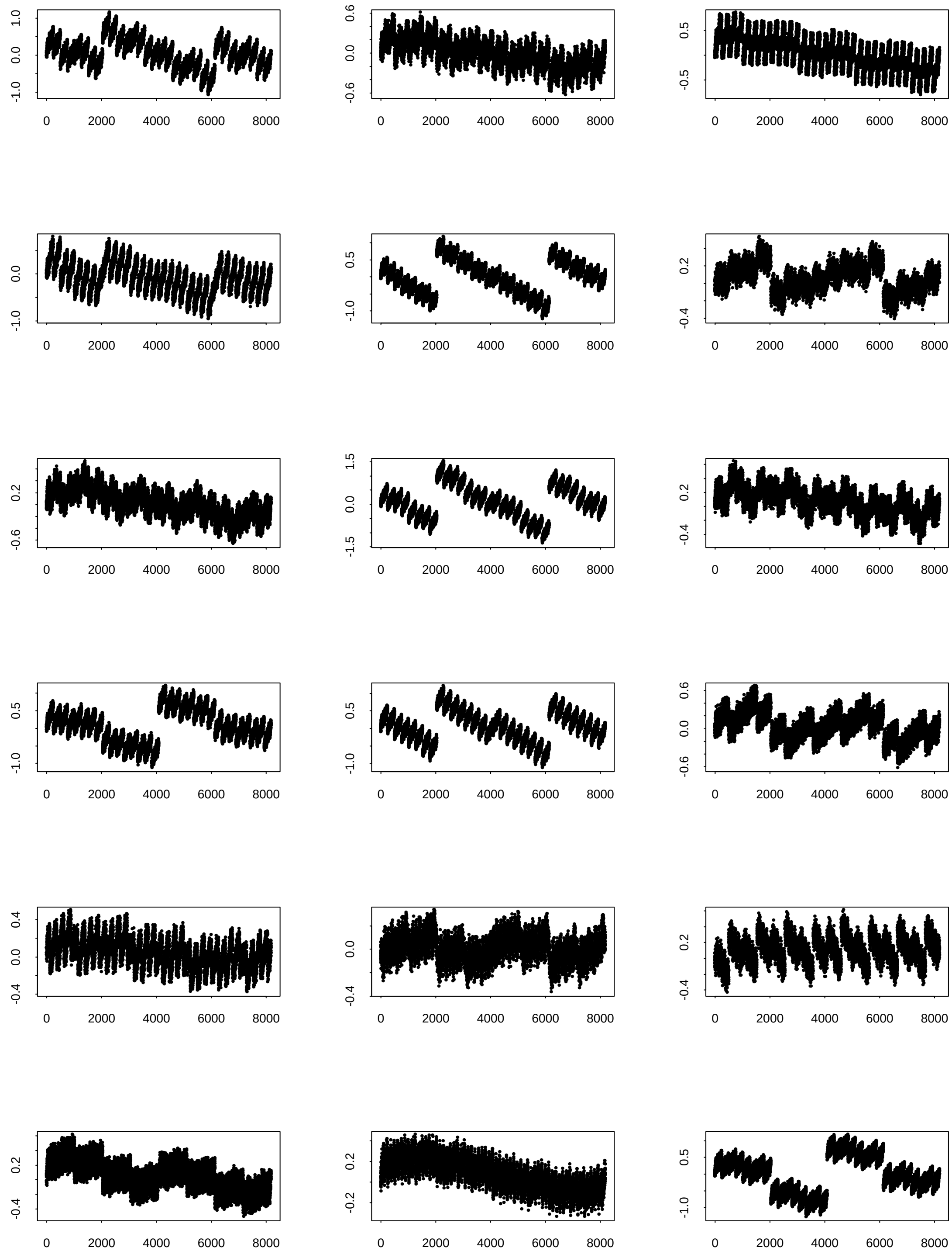


Figure 28: This figure shows 18 of the 88 vectors we analyzed, each corresponding to the 8192 error transition levels of a 13 bit converter.

3.4.4. The Joint Distribution of a Sample Mean and an Extremum, With Applications to Quality Control

Mark G. Vangel
Statistical Engineering Division, ITL

In some industrial applications one compares a sample mean and minimum, or a mean and maximum, to reference values, and determines if the lot from which the sample was taken is acceptable, or if further investigation of this lot is indicated. For example, emphasis has long been placed on checking sample means and minima of lots of various packaged goods. Also, the sample means and minima are used in the testing of batches of raw material by many manufacturers of composite materials. And means and maxima of power loss of sampled motors have been recently proposed for use in testing whether manufactured motors comply with labeled motor efficiencies.

Because the exact joint distribution of an extremum and the mean of a sample is usually complicated, establishing these reference values using statistical considerations typically involves crude approximations or simulation, even under the assumption of normality. Saddlepoint approximations, however, can be used to develop fairly simple and very accurate approximations to the joint cdf of the mean and an extremum.

Let $\phi(t) = \Phi'(t)$ denote the normal density, and let

$$h(t) = \frac{\phi(t)}{1 - \Phi(t)}$$

be the normal hazard function. For a normal model, the joint cdf of the mean and the minimum is well approximated by

$$F_{X_{(1)},\bar{X}}(t_1, t_2) = \frac{\int_{-\infty}^{t_*} \Phi(\sqrt{n}t_2)A(t)dt + \int_{t_*}^{\infty} \Phi\left\{\sqrt{n}\left[t_1 + \frac{n-1}{n}(h(t) - t)\right]\right\} A(t)dt}{\int_{-\infty}^{\infty} A(t)dt},$$

where

$$\begin{aligned} A(t) &= h^{-(n-1)}(t)\exp\left[\frac{(n-1)^2}{2n}(h(t)-t)^2 + (n-1)t(h(t)-t)\right] \\ &\quad \cdot\sqrt{1 - h^2(t) + th(t)}, \end{aligned}$$

and where t_* is the (unique) solution to the equation

$$\frac{n-1}{n}(h(t_*) - t_*) = t_2 - t_1.$$

In the figure, this approximation is compared with the contours of the exact distribution for a sample size of 2. A corresponding approximation has also been obtained for an exponential model.

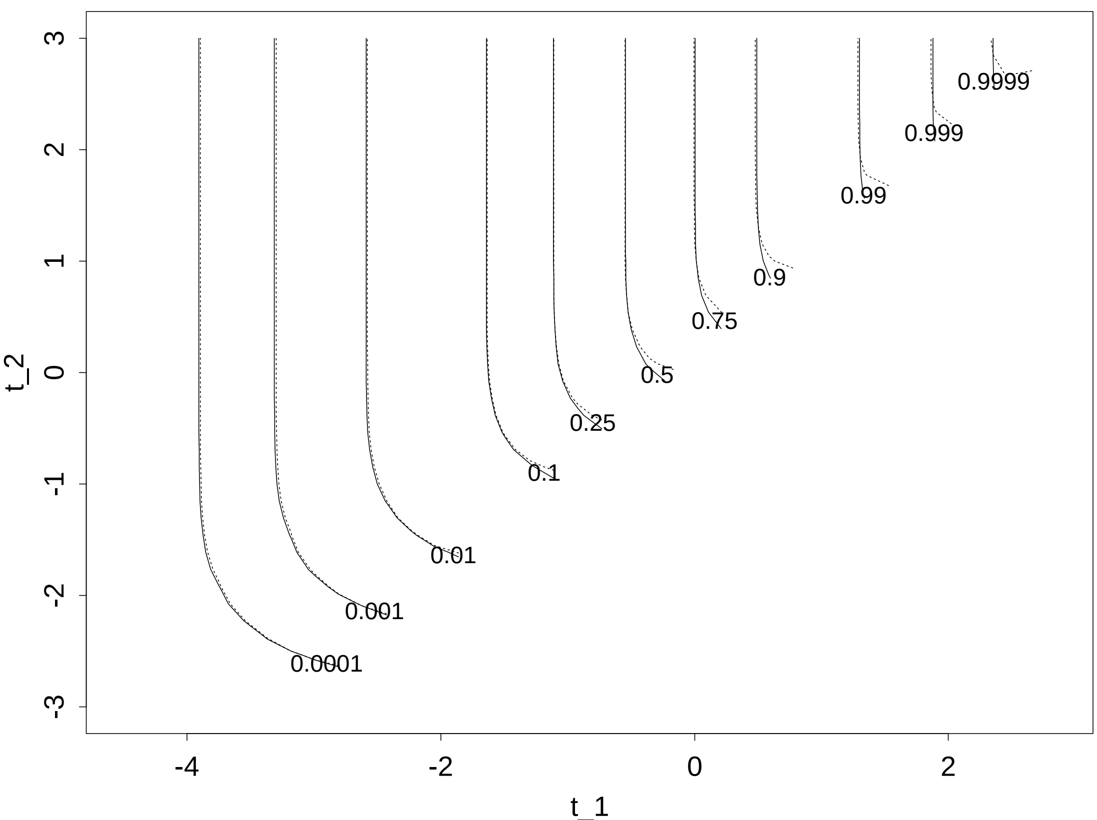


Figure 29: Contours of the joint distribution of the mean and the minimum (broken lines), along with the saddlepoint approximation (solid lines), for a sample size of 2 from a normal population. The approximation is already very accurate, and this accuracy improves rapidly with increasing sample size.

3.4.5. Inference on a Common Mean in an Interlaboratory Study

Mark G. Vangel
Andrew L. Rukhin
Statistical Engineering Division, ITL

Data on a quantity measured by several laboratories often exhibits non-negligible between-laboratory variability, as well as different within-laboratory variances. Also, the number of measurements made at each laboratory can differ. A question of fundamental importance in the analysis of such data is how to best estimate a consensus mean, and what uncertainty to attach to this estimate.

We have been engaged in a detailed investigation of this problem, and its generalizations and applications. Recent results include a representation of the posterior distribution of the common mean under a Bayesian hierarchical model with 'noninformative' prior distributions as a product of symmetric 'generalized t-densities':

$$\pi(\mu, \sigma|\{x_{ij}\}) \propto \pi(\sigma) \prod_{i=1}^{k} \frac{1}{t_i} f_{\nu_i} \left[\frac{x_i - \mu}{t_i}; \frac{2\sigma^2}{t_i^2} \right],$$

where

$$f_\nu(x;\theta) = \frac{1}{\Gamma_{\nu/2}\sqrt{\pi}} \int_0^\infty \frac{y^{(\nu+1)/2-1} e^{-y\left[1+\frac{x^2}{\theta y+\nu}\right]}}{\sqrt{\theta y + \nu}} dy,$$

μ is the common mean, σ is the between-lab standard deviation, x_i is the sample mean, t_i is the sample standard deviation of this mean, and ν_i is the degrees of freedom for this standard deviation.

This posterior distribution can lead to approximate confidence regions for μ and σ for situations where exact frequentist results are not available. The figure illustrates the results of a small simulation study which compares frequentist intervals with non-informative-prior Bayesian intervals for a special case.

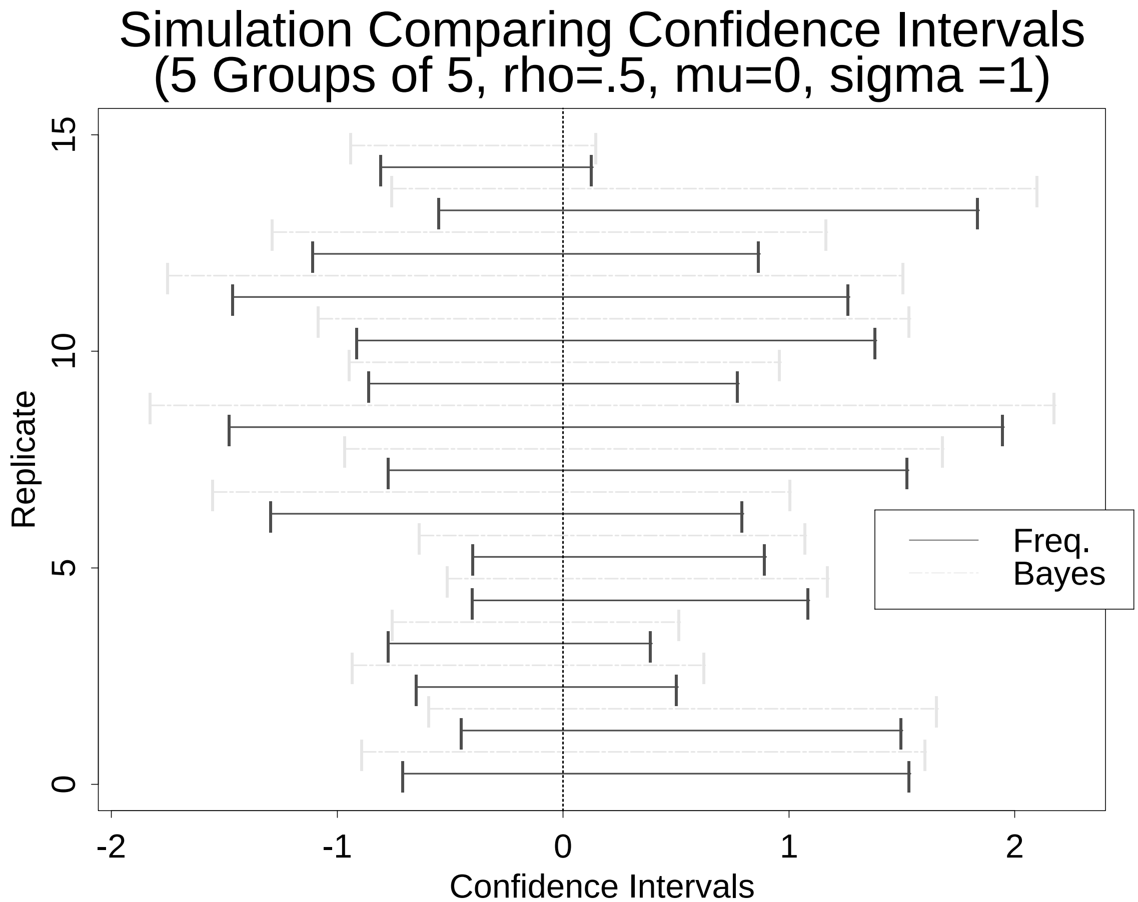


Figure 30: Results of a small simulation comparing Bayesian equal-tailed probability intervals with corresponding confidence intervals on μ for a one-way random-effects ANOVA model with 5 groups of 5, a true mean of 0, standard deviation 1, and intraclass correlation of 1/2. The Bayesian intervals are always slightly larger, which is to be expected since these intervals do not require the assumption that the within-laboratory variances are equal.

3.4.6. Estimating Process Capability Indices for Autocorrelated Processes

Nien Fan Zhang
Statistical Engineering Division, ITL

Process capability indices (CPI) have been widely used in manufacturing industries to measure a process' performance in meeting preset specification limits. They are also used by supplier companies to demonstrate the quality of their products. Among all the capability indices, Cp and Cpk are the most widely used. In the past years there have been a lot of discussions and debates about the use of the process capability indices. Interval estimation of the process capability indices was proposed. In practice, there is also a concern about the assumption of the mutual independence of the process observations. It is well known that in practice process data are often autocorrelated. This is especially true for continuous manufacturing processes such as chemical processes. When the sampling frequency is not too low, the observations are often autocorrelated. In process industries, it is common for quality personnel and process operators to use the capability indices to monitor the process performance. In this case, the variances of the sample CPI's when the data are autocorrelated are needed to construct the interval estimates of CPI.

We assume that the process is a discrete weakly stationary process. Cp and Cpk are defined in the same way as when the process observations are independent. Under the above assumption, the expectation and variance of the sample process variance were derived. It also has been shown that the covariance between the sample process mean and sample process variance is zero when the process is weakly stationary.

Approximate variances of Cp, one-sided Cpk, and Cpk have been derived in similar forms when the process observations are independent. These variances can be easily calculated based on the corresponding CPI, sample size, the process variance and autocorrelations. Thus, the interval estimators of capability indices can be constructed when the process is stationary. In particular, when the process is a first order autoregressive (AR(1)) process, the approximate variances are expressed by the process parameter phi, sample size, process variance and CPI. For a fixed process parameter, The attached figure shows that when the sample size increases, the variance of Cp and Cpk decrease. In the figure, the curves with markers of "o", "*", and "+" are corresponding to the AR(1) processes with phi=0.25, 0.50 and 0.75 respectively.

Simulations have been done to find the coverage probability of k-sigma intervals of Cp and Cpk. The results show that the true Cp and Cpk lie within the interval roughly 99% of the times when k=3 and about 93% of the times when k=2.

This work will be published in *Journal of Applied Statistics.*

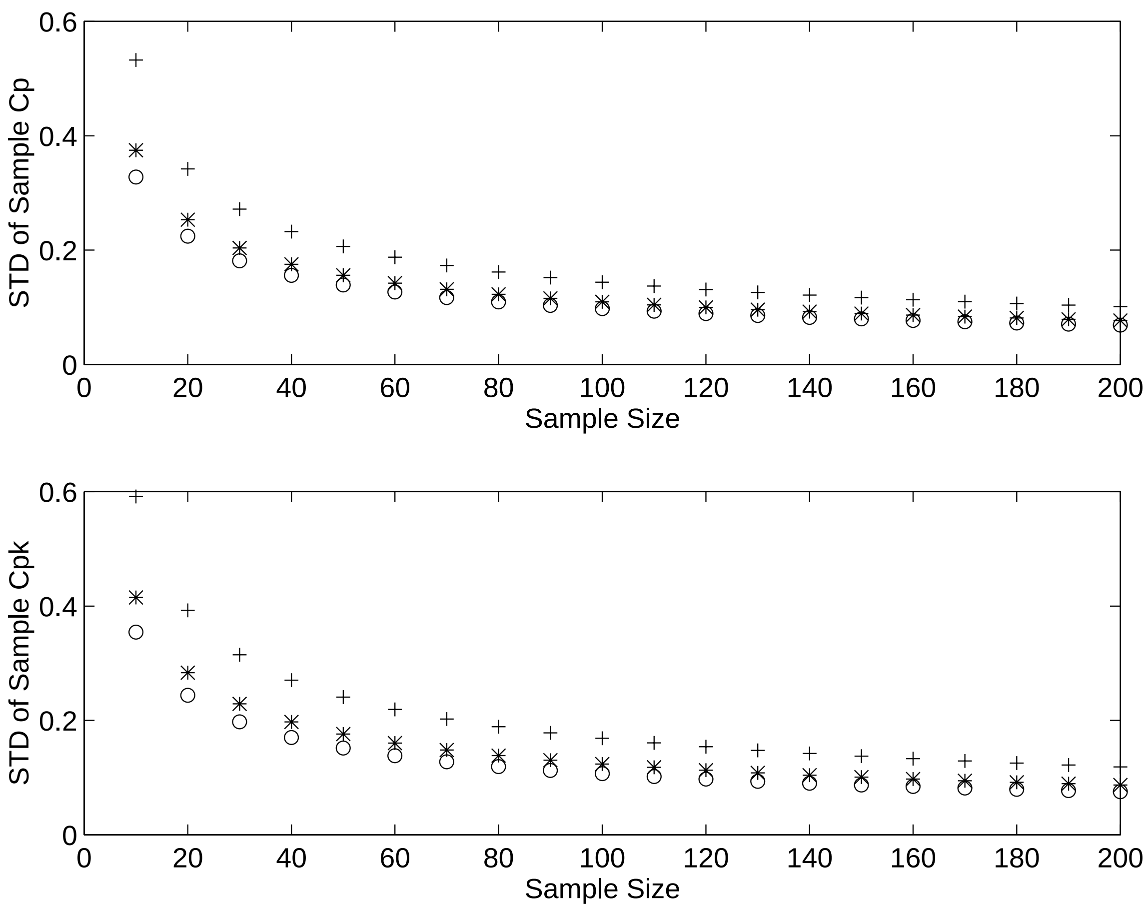


Figure 31: This figure shows that when the sample size increases, the variances of Cp and Cpk decrease.

3.4.7. A Statistical Control Chart for Stationary Process Data

Nien Fan Zhang
Statistical Engineering Division, ITL

Statistical process control (SPC) techniques are widely used in industry for process monitoring and quality improvement. Traditional SPC methodology is based on a fundamental assumption that process data are statistically independent. Process data, however, are not always statistically independent from each other. In the continuous industries such as the chemical industry, most process data are autocorrelated. Under such conditions, traditional SPC procedures are not effective and appropriate for monitoring, controlling, and improving process quality. To accommodate autocorrelated data, some SPC methodologies have been developed in recent years. One approach is to use a process residual chart. This procedure requires one to model the process data and obtain the process residuals. Assuming a true model, the residuals are statistically uncorrelated to each other. Then, traditional SPC charts such as X charts (Shewhart individual charts), CUSUM charts, and exponentially weighted moving average (EWMA) charts can be applied to the residuals.

Use of a residual chart has the advantage that it can be applied to any autocorrelated data even if the data are from nonstationary processes. However, the residual charts do not have the same properties as the traditional charts. In addition, time series modeling is often awkward in the SPC environment. Although automatic-modeling algorithms can be used to obtain the process residuals, this approach often requires much effort in practice. The user of a residual chart must check the validity of the model over time to reduce the mixed effects of modeling error and process change.

In this article, I propose a new SPC chart, the EWMAST chart, for stationary process data. The chart is constructed by charting the traditional EWMA statistic. The control limits are based on the approximate standard deviation of the EWMA, which has been analytically derived. This EWMAST chart does not require any time series modeling effort. I compare the EWMAST chart with the residual X chart and other charts via the average run length. Simulation study shows that the EWMAST chart performs better than the residual X chart, X chart, and other charts when the process autocorrelation is not very positively strong and the mean shifts are small to medium. In the accompanying figure, ARL of EWMAST chart with parameter 0.2 and the residual chart for AR(1) processes with phi = 0.25, 0.5, 0.75, and 0.95 are plotted when the mean shifts are 0, 0.5, 1, 2, and 3 in the unit of process standard deviation. Solid lines with circles indicate the ARL of EWMAST charts while dot lines with asterisks indicate the ARL of the residual charts. The ARL are on a logarithmic scale with base = 10. For a very wide range of autocorrelated data including uncorrelated data as a special case, I recommend to use the EWMAST chart with 3-sigma control limits and parameter 0.2 to monitor the process mean.

This paper is published in *Technometrics* (1998), 40, 24-38.

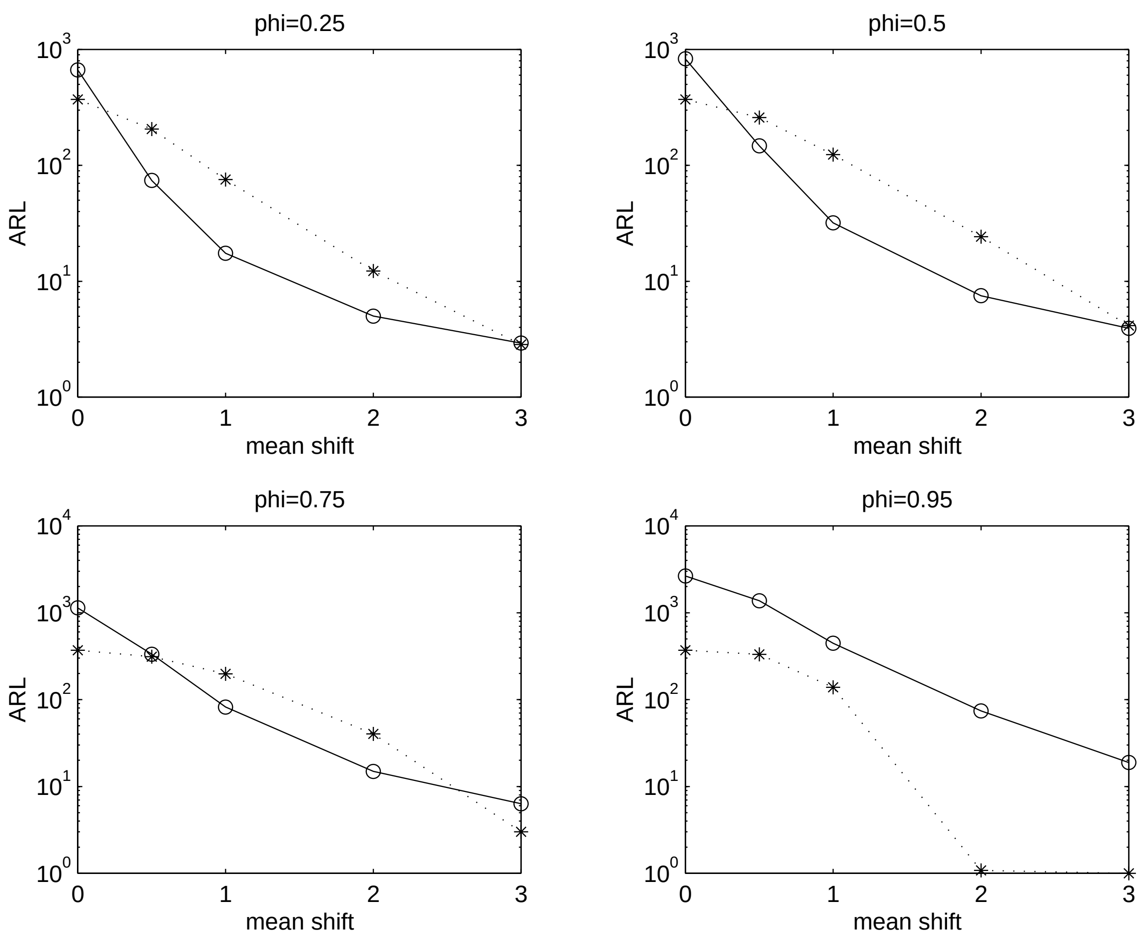


Figure 32: This figure shows the ARLs of the EWMAST and residual charts applied to AR(1) processes.

4. Conferences, Workshops and Seminars.

4.1. Annual Meeting of the Classification Society of North America

David L. Banks
Statistical Engineering Division, ITL

The forty-first annual meeting of the Classification Society of North America was held on June 12-15, 1997 at American University. Rob Tibshirani of the University of Toronto and John Hartigan of Yale University were the plenary speakers; Joseph Kruskal of AT&T Laboratories gave the banquet address. Special topic sessions included Public Health Statistics, Phylogenetic Inference, DNA Fingerprinting, Environmental Applications, Issues in Computer Security, Clustering Methods, Applications in Imaging, and Strategies for Neural Nets. The conference was organized by David Banks of NIST.

4.2. Symposium on Visualization and Testing in Intrusion Detection

David L. Banks
Mark Levenson
Statistical Engineering Division, ITL

Joseph Konczal
Sharon Laskowski
Information Access and User Interfaces Division, ITL

Donald Marks
Computer Security Division, ITL

A symposium on issues and ideas in the use of statistical visualization methods for network intrusion detection was held on November 24, 1997 at NIST. The invited speakers were Roy Maxion, Carnegie Mellon University; Gene Spafford and Carla Brodley of Purdue University; K. Y. Tan of the University of Melbourne; David Banks, Mark Levenson, and Sharon Laskowski of NIST; Georges Grinstein of the University of Massachusetts at Lowell; and James Codespotte of the NSA. There were 74 attendees, drawn from government, industry, the armed forces, and academics. The symposium was organized by David Banks of NIST.

4.3. Easy and Not-So-Easy Methods of Uncertainty Analysis

Carroll Croarkin
Mark Levenson
Jack Wang
Statistical Engineering Division, ITL

Theodore Doiron
Precision Engineering Division, MEL

John Wehrmeyer
Eastman Kodak

Concepts of uncertainty analysis for metrology were outlined in a two-day workshop presented in conjunction with the Measurement Science Conference in Pasadena, CA on Feb. 2-3, 1998.

Croarkin began the workshop with a discussion of the ISO policy on uncertainty and a case study that illustrated the the estimation of type A components of uncertainty from multiple sources, expanded uncertainties and effective degrees of freedom.

Doiron discussed his experience and reliance on check standards for computing uncertainties for dimensional measurements.

Wehrmeyer provided an industrial persepective with suggestions for how 'easier' assessments of uncertainty can sometimes be inferred from guard-banding techniques and other sources.

Wang illustrated how an uncertainty assessment can be made from an inter-laboratory study, in this case, for measurements on 10 volt zener standards.

In his first talk, Levenson outlined a method for computing uncertainties of results from a linear calibration line, and in a second talk, he showed a case study of how formal propagation or error was used to compute uncertainties for radiometry measurements.

Croarkin concluded with a discussion of pros and cons of propagation of error and an uncertainty budget that combined results from type A evaluations with type B evaluations with propagation of error.

4.4. Advanced Mass Measurements

Carroll Croarkin
Statistical Engineering Division, ITL

Georgia Harris
Office of Weights and Measures, Technology Services

A 5-day workshop on mass measurements was given at NIST in March 1997 for industrial and state metrologists, weight manufacturers and balance manufacturers. Participants were required to have taken two lower level courses before registering for this workshop which, in large part, focuses on statistical methods for controlling mass measurement processes and assessing uncertainties of mass determinations.

Attendees were introduced to concepts of: theory and solution of weighing designs; error analysis; components of variance; and procedures for controlling measurement processes. Specific applications to mass measurements included:weighing designs for common weight sets; propagation of uncertainties through several series of designs; computations of uncertainty from check standard measurements; control of balance precision; and computation of total uncertainties. The course materials were supplemented by interactive computerized workshops which strengthen skills in the areas of weighing designs and process control.

4.5. Statistical Methods for Mass Metrology

Carroll Croarkin
Statistical Engineering Division, ITL

Jerry Keller
Automated Technology Production Division, Manufacturing Engineering Laboratory

A workshop on statistical treatment of mass measurements was given at the Technological University of Panama, Panama City, Panama in May 1997. The workshop was sponsored by the Organization of American States and the Ministry for Technology of Panama to support the System for Inter-American Metrology where NIST is the lead laboratory for mass mesurements.

Participants from the national laboratories of Central America, who are already proficient in mass metrology, were trained in statistical procedures relating to: theory and solution of weighing designs, propagation of uncertainties through several series of weighing designs; statistical control of the measurement process using check standards; precision of the balances; and computation of final uncertainties.

4.6. International Standards Activities

Carroll Croarkin
Statistical Engineering Division, ITL

The Statistical Engineering Division supports the development of international standards, particularly those that impact measurement science, via participation on ISO Technical Committee 69 on Statistical Methods and sub-committee SC 6 on Measurement Methods. Clients for TC 69 standards are US industry and other technical committees of ISO which produce standards with statistical content.

Currently, a document on Capability of Detection, which outlines statistical methods for quantifying the smallest amout of a trace element that can be detected by an instrument, is under development. A document on Specification Limits is also under development. It has twin goals of outlining methods for the producer to set specification limts and methods for the consumer to test product for conformance against the agreed upon limits.

Croarkin is convenor of a new working group, ISO/TC69/SC6/WG7, which will report on the relationship between the ISO Guide to the Expression of Uncertainty in Measurement (GUM) and ISO 5752 - Accuracy of Measurement Methods and Results. This report is expected to provde the framework for SC 6 to develop a document, or series of documents, on statistical methods to be used in conjunction with the GUM to assess uncertainties associated with measurement results.

4.7. Statistics for Scientists and Engineers: A Program of Short Courses and Tutorials

Mark Vangel, David Banks, Jim Filliben, Will Guthrie, Eric Lagergren, Stefan Leigh, Mark Levenson, Nien Fan Zhang
Statistical Engineering Division, ITL

Statistical concepts and methods are indispensable to research efficiency and planning as well as the characterization of uncertainty in measurements. Hence, the Statistical Engineering Division offers a program of short courses (and shorter tutorials), primarily for the NIST community. The main objective of this effort is to develop an appreciation of the meaning and usefulness of basic statistical concepts and techniques, leading at least to the ability to interpret reliably statistical analyses performed by others. In addition, sufficiently motivated scientists or engineers will be able to learn to perform their own basic statistical analyses. Courses and Tutorials presented during 1997, or planned for 1998, are:

- **Time Series Analysis**
  Nien Fan Zhang

1/7/97, 1/14/97, 1/21/97, 9-12

- **Spectral Methods for Data Analysis**
  Mark S. Levenson
  6/24/97, 9-12
- **Exploratory Data Analysis**
  James J. Filliben
  11/14/97, 11/18/97, 11/25/97, 12/9/97, 12/16/97, 9-12
- **Analysis of Variance**
  Stefan Leigh
  1/14/98, 1/21/98, 1/28/98, 2/4/98, 2/11/98, 9-4
- **Multivariate Nonparametric Regression**
  David Banks
  1/12/98, 1/13/98, 10-2
- **Experiment Design** Eric Lagergren
  4/9/98, 4/16/98, 4/23/98, 4/30/98, 9-12
- **Statistical Methods for Assessing Measurement Uncertainty**
  Mark Levenson and Mark Vangel
  6/4/98, 6/11/98, 6/18/98, 6/25/98, 7/2/98, 2-5
- **Regression Models**
  William Guthrie
  10/13/98, 10/15/98, 10/20/98, 10/22/98, 10/29/98

5. Special Programs

5.1. Standard Reference Materials

Carroll Croarkin
Statistical Engineering Division, ITL

The Statistical Engineering Division supports the Standard Reference Materials Program and the other NIST laboratories by collaborating directly with chemists and other scientists engaged in the certification of Standard Reference Materials (SRMs). SRMs are artifacts or chemical compositions that are manufactured according to strict specifications and certified by NIST for one or more quantities of interest. SRMs are a primary vehicle for disseminating measurement technology to industry.

Development of a new SRM typically takes about five years and encompasses: 1) validation of a measurement method; 2) design of a prototype; 3) stability testing; 4) study of measurement error; 5) certification uncertainty analysis. Statisticians advise on the design and analysis of experiments at all phases; develop estimation methods for data from different analytical methods; help reconcile interlaboratory differences; and combine all information to produce a certified value and statement of uncertainty.

In 1997, thirty-five new SRMs were assigned to division staff; fifty SRM certifications were completed; and approximately fifty SRMs are still in some phase of development. The largest number of SRMs, far and away, came from the Chemical Science Technology Laboratory (CSTL), but SRMs from the other NIST laboratories covered a variety of applications, for example: MEL (e.g. sinusoidal roughness); MSEL (e.g. Knoop hardness); EEEL (e.g. resistivity of silicon wafers) PL (e.g. optical density filters); BFRL (e.g. thermal resistance of fibrous glass insulation).

Tracking such a large number of SRMs has always been challenging because the responsibility for the status of the certification process often shifts back and forth between scientist and statistician as follow-on data are taken and analyzed or more analytes are added to the certification process. This year an internal web page was put on-line that allows NIST staff to determine the status of SRMS in SED.

The large workload has led the division to consider more efficient methods for handling the statistical design and analyses of SRMs. As a start, a common protocol for certifying gas mixture standards was developed in collaboration with chemists from CSTL. From this protocol, chemists developed a spread-sheet for handling certifications and now require only occasional assistance from statisticians for fifty or more issues per year.

5.2. On-line Engineering Statistics Handbook

Carroll Croarkin, James J. Filliben, William F. Guthrie, Alan Heckert
Statistical Engineering Division, ITL

Paul Tobias, Jack Prins, Chelli Zey
SEMATECH

Barry Hembree
AMD

Patrick Spagon
Motorola University

The Statistical Engineering Division is pursuing a joint project with the Statistical Methods Group of SEMATECH to develop and publish an on-line handbook on statistical methods for scientists and engineers which provides modern statistical solutions to problems confronting U.S. industry, particularly the semiconductor industry, and the NIST laboratories. The graphic on the facing page shows the first level outline for the handbook.

The initial inspiration for the project started with the idea of updating NBS Handbook 91, *Experimental Statistics* by Mary Natrella, and evolved to a document for browsing on the World Wide Web complete with supporting software.

The public domain software, Dataplot, developed at NIST by J. Filliben, is currently interfaced with the handbook. This software is coupled with detailed case studies that lead the reader through statistical approaches to scientific problems and encourage analyses of their own data.

All chapters are under development and some sections are available for internal review and comment. Usability testing protocols are also under development with the assistance of the Vizualization and Virtual Reality Group of the ITL.

During the past year, layout and navigation for 2000 pages of html text were put in place, and demonstrations of chapter materials and software were given for the managements of AMD, Motorola University and for the SEMATECH Advisory Council.

In August, 1997, the handbook was featured at an invited poster session sponsored by the American Statistical Association at the Joint Statistical Meetings. In March 1998, the handbook will be demonstrated to engineers at the Ultra Large Scale Integrated Circuit Conference at NIST, and data on usability will be collected from volunteers.

In the coming year, all chapters will be completed; analysis macros for specialized case studies will be developed; the handbook will be interfaced with other statistical software packages; and a peer review process will begin. The handbook is supported by the Systems Integration for Manufacturing project.

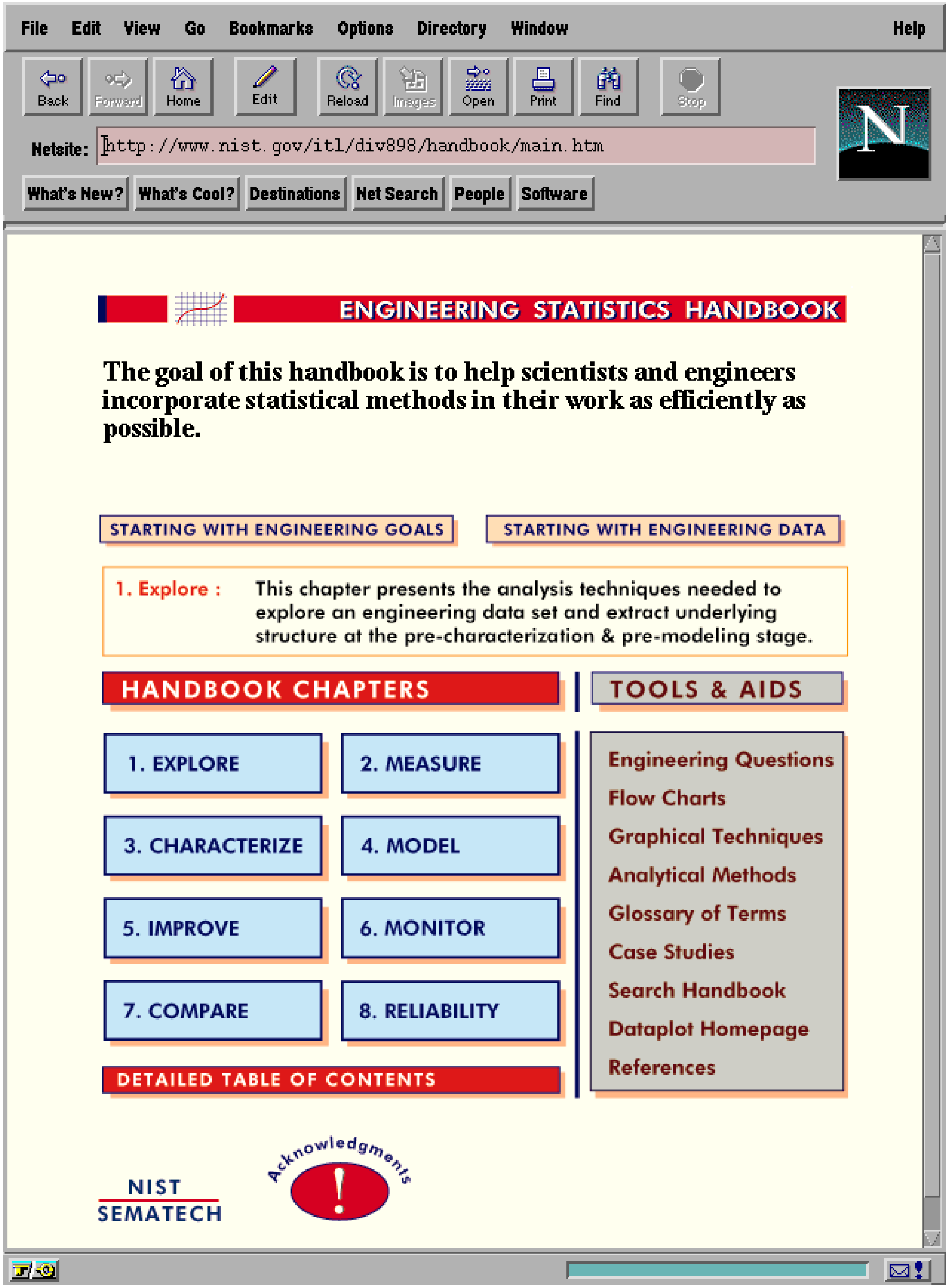


Figure 1: The main page of the handbook showing entry points to the text

5.3. Statistical Reference Datasets

Eric S. Lagergren, M. Carroll Croarkin, James J. Filliben, Lisa M. Gill, William F. Guthrie, Hung-kung Liu, Mark G. Vangel, Nien Fan Zhang
Statistical Engineering Division, ITL

Janet E. Rogers, Bert W. Rust
Mathematical & Computational Sciences Division, ITL

Phoebe Fagan
Standard Reference Data Program, TS

With the widespread use and availability of statistical software, concerns about the numerical accuracy of such software are now greater than ever. Inevitably, numerical accuracy problems can exist with some of this software despite extensive testing. Indeed, this has been a continuing cause of concern for statisticians. Many have cited the need for an easily-accessible repository of reference datasets. To date no such collection has been available. In response to concerns of both the statistical community and industrial users, the Statistical Engineering Division in collaboration with the Mathematical & Computational Sciences Division and Standard Reference Data Program have developed a Web-based service that provides reference datasets with certified values for a variety of statistical methods. This service is called Statistical Reference Datasets (StRD).

Currently 58 datasets with certified values are provided for assessing the accuracy of software for univariate statistics, analysis of variance, linear regression, and nonlinear regression. The collection includes both generated and "real-world" data of varying levels of difficulty. Generated datasets are designed to challenge specific computations. These include the classic Wampler datasets for testing linear regression algorithms and the Simon & Lesage datasets for testing analysis of variance algorithms. Real-world data include challenging datasets such as the Longley data for linear regression, and more benign datasets such as the Daniel & Wood data for nonlinear regression.

Certified results for linear procedures were obtained using extended precision software to code simple algorithms for each type of computation. Carrying 500 digits through all of the computations allowed calculation of output unaffected by floating point representation errors. Certified values for nonlinear regression are the "best-available" solutions, obtained using 64-bit precision and confirmed by at least two different algorithms and software packages using analytic derivatives.

The team officially released the StRD web service in August 1997 and spent the latter part of the year publicizing the web service. A special contributed paper session was presented at the 1997 Joint Statistical Meetings in August. Talks were also given at NIST Gaithersburg and Boulder. The StRD home page has been hit approximately 900 times each month. In the coming year, we plan to publish a NIST Technical Report documenting the development of the StRD web service and collect feedback from users as to how to improve the web service.

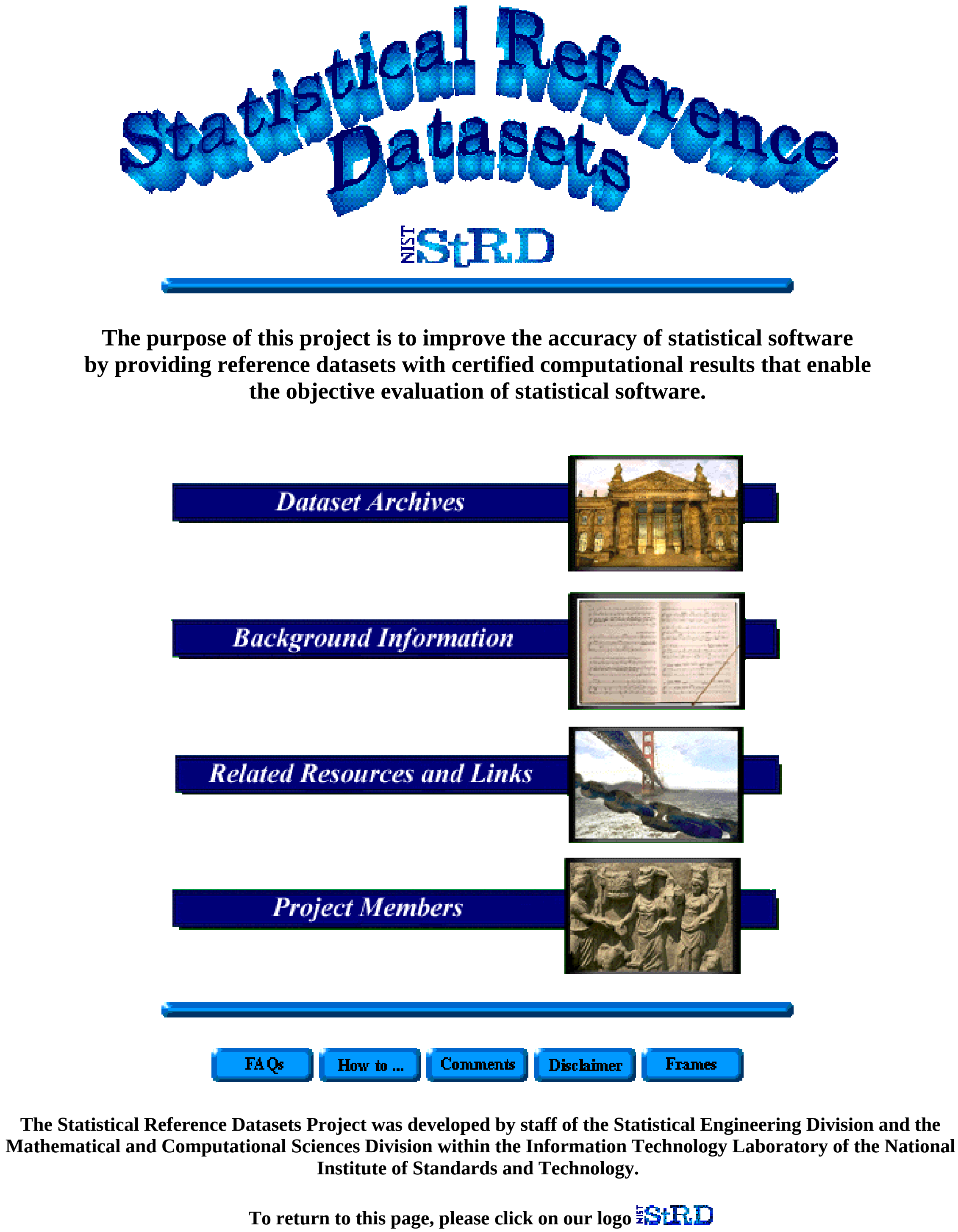


Figure 2: The StRD home page.

5.4. MIL-HDBK-17: Composite Materials Handbook

Mark Vangel
Statistical Engineering Division, ITL

Mark Vangel is Chairman of the Statistics Working Group of Mil-Handbook-17, which develops and publishes statistical methods for composite materials. These materials, which can have exceptionally high ratios of strength and stiffness to weight, are of growing importance, particularly in the aerospace industry. However, strength properties of composite materials typically exhibit considerable variability, due to the brittleness of most fibers and many matrices, and due to processing. Statistical methods (specifically methods for tolerance limits, mixed model analysis, and quality control) are thus important to the use of these materials. Mil-Handbook-17 is an evoloving document which is intended to be used as a primary reference, both for data and for guidelines on data analysis, by composites engineers, by the Department of Defense, and by regulatory agencies.

6. Staff Publications and Professional Activities

6.1. Publications

6.1.1. Publications in Print

1. D.L. Banks (co-editor with S. Kotz and C.B. Read). *Encyclopedia of Statistical Sciences, Update Volume 1*, 1997.
2. D.L. Banks (co-editor with S. Kotz and C.B. Read). *Encyclopedia of Statistical Sciences, Update Volume 2*, 1998.
3. D.L. Banks (with T. Kurfess and L. Wolfson) A multivariate statistical approach to metrology, *Journal of Manufacturing Science and Engineering*, 118, 1996, p. 652-658.
4. D.L. Banks (with W. Shannon) An MLE strategy for combining CART models, *Proceeding of the 1997 Symposium on the Interface: Computing Science and Statistics*, p. 227-232.
5. D.L. Banks (with M. Scher, G. Dokianakis, D. Steppe, and R. Sclabassi) Computer classification of state in healthy preterm neonates, *Sleep*, 20, 1997, 132-141.
6. K.J. Coakley, Nonequilibrium Kinetics of Neutral Atoms Confined in a Harmonic Potential, *PHYSICA A*, 234,1996, pp.407-426
7. K.J. Coakley with (M.Holland,J.Williams, and J.Cooper), Trajectory Simulation of Kinetic Equations for Classical Systems, *Quantum Semiclassical Optics*, 8, 1996, pp.571-581
8. K.J. Coakley, A Bootstrap Method for Nonlinear Filtering of EM-ML Reconstructions of PET Images, *International Journal of Imaging Science and Technology*, 7(1), 1996, pp.54-61
9. K.J. Coakley and S. Wilson, Analysis of Asymmetry in Physics, *Physical Review E*, 53(4), 1996, pp. 2160-2168
10. K.J. Coakley (with S.S.-C. Tai, R.G. Christiansen, P. Ellerbe, T. Long, M. Welch) The Certification of Phencyclidine in Lyophilized Human Urine Reference Materials, *Journal of Analytical Chemistry*, 20, 1996, pp. 43-49
11. K. Ehara, C.Hagwood and K.J.Coakley, Novel Method to Classify Aerosol Particles According to Their Mass-to-Charge Ratio," *Journal of Aerosol Science* , Vol. 27, No. 2, pp.217-234, 1996
12. C. Croarkin (with T.V. Vorburger, J. F. Song, C.H.W. Giauque, T.B. Renegar,

E.P. Whitenton, Stylus-laser Surface Calibration System, *Precision Engineering*, 1996, p. 157-163.

13. J.J. Filliben, (with K.C. Li) A systematic approach to the analysis of complex interaction patterns in two-level factorial design, *Technometrics*, 39 (3), 1997, pp. 286–297.
14. L.M. Gill, J.J. Filliben, (with S. Jahanmir, L.K. Ives), Effect of Grinding on Strength of Sintered Reaction Bonded Silicon Nitride, *2nd International Conference on Machining of Advanced Materials (MAM)*, VDI Berichte 1276, (1996) pp.603-615.
15. W.F. Guthrie (with M.W. Cresswell, R.A. Allen, L.W. Linholm, W.B. Penzes and A.W. Gurnell), Hybrid Optical-Electrical Overlay Test Structures, *IEEE Transactions on Semiconductor Manufacturing*, 10 (2), 1997.
16. W.F. Guthrie (with N.J. Carino and G.M. Mullings), Evaluation of ASTM Standard Consolidation Requirements for Preparing High-Strength Concrete *High Performance Concrete: Design and Materials and Recent Advances in Concrete Technology, ACI SP-172, Proc. of the Third ACI International Conf. on High-Performance Concrete*, Kuala Lumpur, Malaysia, December 2-5, 1997, pp 733-768.
17. W.F. Guthrie (with W.E. Lee, M.W. Cresswell, R.A. Allen, J.J. Sniegowski, and L.W. Linholm), Reference-Length Shortening by Kelvin Voltage Taps in Linewidth Test Structures Replicated in Monocrystalline Silicon Films, *Proc. of IEEE International Conf. on Microelectronic Test Structures*, Monterey, CA, March 17-28, 1997.
18. W.F. Guthrie (with M.W. Cresswell, J.J. Sniegowski, R.N. Ghoshtagore, R.A. Allen and L.W. Linholm), Electrical Linewidth Test Structures Fabricated in Monocrystalline Films for Reference Material Applications, *Proc. of IEEE International Conf. on Microelectronic Test Structures*, Monterey, CA, March 17-28, 1997.
19. W.F. Guthrie (with M.W. Cresswell, J.J. Sniegowski, R.N. Ghoshtagore, R.A. Allen, A.W. Gurnell, L.W. Linholm, R.G. Dixson, and E.C. Teague), Recent Developments in Electrical Linewidth and Overlay Metrology for Integrated Circuit Fabrication Processes, *Japanese Journal of Applied Physics* December 1996, Vol. 35, Part 1, No. 12B, pp. 6597-6609.
20. W.F. Guthrie (with M.W. Cresswell, R.A. Allen, L.W. Linholm, and A.W. Gurnell), Hybrid Optical-Electrical Overlay Test Structure, *Proc. of IEEE International Conf. on Microelectronic Test Structures*, Trento, Italy, March 25-28, 1996.
21. W.F. Guthrie (with M.A. Shen, F. Mopsik, W. Wu, W.E. Wallace, N.C. Beck Tan, and G.T. Davis), Advances in the Measurement of Polymer CTE: Micrometer to Atomic-Scale Measurements, *Proceedings of the 211th American Chemical Society National Meeting, Division of Polymer Chemistry*, 37(1), New Orleans, LA, March 4, 1996, pp. 180–182.
22. A. Heckert, (with E. Simiu), Extreme Wind Distribution Tails: A "Peaks Over Threshold" Approach, *Journal of Structural Engineering*, May, 1996.
23. M.S. Levenson (with T. Hopp), Performance Measures for Geometric Fitting in the NIST Algorithm Testing and Evaluation Program for Coordinate Measurement Systems, *Journal of Research of the National Institute of Standards and Technology*, Vol 100, No 5, 1995, pp. 563–574.

24. M.S. Levenson (with D. Bentz, N. Martys, P. Stutzman, E. Garboczi, J. Dunsmuir, L. Schwartz), X-Ray Microtomography of an ASTM C109 Mortar Exposed to Sulfate Attack, *Material Research Society Symposium Proceedings*, Vol 370, 1995.
25. M.S. Levenson, Removing Quantization Noise Using Wavelets, *Proceedings of the Section on Physical and Engineering Sciences of the 1996 Joint Statistical Meetings*, pp. 193–197.
26. M.S. Levenson (with D. Bright, J. Sethuraman), Adaptive Smoothing of Images with Local Weighted Regression, *Statistical and Stochastic Methods for Images Processing, Proceedings of the SPIE Conference*, Vol 2823, 1996, pp. 85–99.
27. M.S. Levenson and K.R. Eberhardt (with W.T. Estler, S.D. Phillips, B. Borchardt, T. Hopp, C. Witzgall, M. McClain, Y. Shen, X. Zhang), Error Compensation for CMM Touch Trigger Probes, *Precision Engineering*, Vol 19, No 2/3, 1996, pp. 85–97.
28. M.S. Levenson, An Overview of Statistical Uncertainty Analysis, *NIST Special Publication on the 1996 Measurement Quality Conference.*
29. M.G. Vangel and M.S. Levenson (with M. Behlke, R. Saraswati, E. Mackey, R. Demiralp, B. Porter, V. Mandic, S. Azemard, M. Horvat, K. May, H. Emons, S. Wise), Certification of Three Mussel Tissue Standard Reference Materials (SRMs) for Methylmercury and Total Mercury Content, *Fresenius' Journal of Analytic Chemistry*, Vol 358, No 3, 1997, pp. 431–440.
30. M.S. Levenson and M.G. Vangel (with R.E. Scholten, R. Gupta, J.J. McClelland, R.J. Celotta), Laser Collimation of a Chromium Beam, *Physics Review A*, Vol 55, No 2, 1997, pp. 1331-1338.
31. M.S. Levenson and K.R. Eberhardt (with S.D. Phillips, B. Borchardt, D. Sawyer, W.T. Estler, D. Ward, M. McClain, B. Melvin, T. Hopp, Y. Shen), The Calculation of CMM Measurement Uncertainty Via the Method of Simulation by Constraints, *Proceedings of the American Society for Precision Engineering*, Vol 16, 1997, p. 443.
32. K.R. Eberhardt and M.S. Levenson (with R.G. Gann), Fabrics for Testing the Ignition Propensity of Cigarettes, *Fire and Materials*, Vol 21, 1997, pp. 259-61.
33. M.S. Levenson and K.R. Eberhardt (with W.T. Estler, S.D. Phillips, B. Borchardt, T. Hopp, M. McClain, Y. Shen, X. Zhang), Practical Aspects of Touch Trigger Probe Error Compensation, *Precision Engineering*, Vol 21, No 1, 1998, pp. 1–17.
34. M.S. Levenson and K.R. Eberhardt (with S.D. Phillips, W.T. Estler), Calculation of Measurement Uncertainty Using Prior Information, *Journal of Research of the National Institute of Standards and Technology*, to appear.
35. H.K. Liu, (with K. Ehara) Background corrected statistical confidence intervals for particle contamination levels, *Proceedings of the 13th International Symposium on Contamination Control* 478–485, 1996.
36. H.K. Liu, High-dimensional empirical linear prediction, *Advanced Mathematical Tools in Metrology III*, 79–90, 1997.
37. M.G. Vangel, Confidence Limits for a Normal Coefficient of Variation, *The American Statistician*, 50, 1996, pp. 21-26.
38. M.G. Vangel, Design Allowables From Regression Models Using Data From Several Batches, in STP 1274, *Proceedings of the 12th Symposium on Composite Materials: Testing and Design*, American Society for Testing and Materials, 1996.

39. Mark Vangel (with Walter Rossiter, Kevin Kraft and Edward Embree), EPDM Membrane Seam Performance, *Roofing and Building Insulators Contractors Guide*, February, 1998, pp. 65-68.
40. A.L. Rukhin and M.G. Vangel, Estimation of a Common Mean and Weighted Means Statistics, *Journal of the American Statistical Association*, 93, pp. 303-309, 1998.
41. Mark G. Vangel (with Walter J. Rossiter, Kevin M. Kraft and Edward Embree), Performance of Tape-Bonded Seams of EPDM Membranes: The Effect of Load on Peel-Creep, in *Fourth International Symposium on Roofing Technology: Challenges for the 21st Century*, National Roofing Contractors Association, 1997, pp. 1-13.
42. C.M. Wang, H.K. Iyer, Sampling plans for obtaining tolerance intervals in a balanced one-way random-effects model, *Communications in Statistics – Theory and Methods*, 25 (2), 1996, pp. 313–324.
43. C.M. Wang, (with C.T. Lam), Confidence limits for proportion of conformance, *Journal of Quality Technology*, 28 (4), 1996, pp. 439–445.
44. C.M. Wang, (with P.D. Hale, R. Park, W.Y. Lau), A transfer standard for measuring photoreceiver frequency response, *Journal of Lightwave Technology*, 14 (11), 1996, pp. 2457–2466.
45. C.M. Wang, D.F. Vecchia, (with M. Young, N.A. Brilliant), Robust regression applied to optical fiber dimensional quality control, *Technometrics*, 39 (1), 1997, pp. 25–33.
46. C.M. Wang, H.K. Iyer, (with E.B. Brown), Tolerance intervals for assessing individual bioequivalence, *Statistics in Medicine*, 16, 1997, pp. 803–820.
47. C.M. Wang, (with A.H. Rose, S.M. Etzel), Verdet constant dispersion in annealed optical fiber current sensors, *Journal of Lightwave Technology*, 15 (5), 1997, pp. 803–807.
48. C.M. Wang, (with K.B. Rochford, A.H. Rose), NIST study investigates retardance uncertainty, *Laser Focus World*, May, 1997, pp. 223–227.
49. C.M. Wang, (with J.D. Splett), Consensus values and reference values illustrated by the Charpy machine certification program, *Journal of Testing and Evaluation*, 25 (3), 1997, pp. 308–314.
50. C.M. Wang, (with C.T. Lam), A mixed–effects model for the analysis of circular measurements, *Technometrics*, 39 (2), 1997, pp. 119–126.
51. C.M. Wang, (with K.B. Rochford, A.H. Rose, P.A. Williams, I.G. Clarke, P.D. Hale, G.W. Day), Design and performance of a stable linear retarder, *Applied Optics*, 36, (25), 1997, pp. 6458–6465.
52. C.M. Wang, (with P.A. Williams, A.H. Rose), Rotating–polarizer polarimeter for accurate retardance measurement, *Applied Optics*, 36, (25), 1997, pp. 6466–6472.
53. C.M. Wang, (with K.B. Rochford), Accurate interferometric retardance measurements, *Applied Optics*, 36, (25), 1997, pp. 6473–6479.
54. N.F. Zhang, A Statistical Control Chart for Stationary Process Data, *Technometrics*, 40, 1998, pp. 24–38.
55. N.F. Zhang, (with M.T. Postek, R.D. Larrabee), Statistical Models for Estimating the Measurement of Pitch in Metrology Instruments, *Metrologia*, 34, 1997.
56. N.F. Zhang, Detection Capability of Residual Control Chart for Stationary Processes, *Journal of Applied Statistics* , 24(4), 1997, pp. 475–492.

57. N.F. Zhang, Autocorrelation Analysis of Some Linear Transfer Function Models and Its Applications in the Dynamic Process Systems, *Lectures in Applied Mathematics*, 33, 1997, pp. 385–399.
58. N.F. Zhang, (with M.T. Postek, R.D. Larrabee, A.E. Vladar, W.J. Keery, S.N. Jones), A Statistical Measure for the Sharpness of SEM Images, *Proceedings of SPIE*, 1997, vol. 2050, pp. 375–386.
59. N.F. Zhang, Estimating Process Capability Indices for Autocorrelated Processes, *Proceedings of the Section on Quality and Productivity of American Statistical Society*, 1996, pp. 49–54.
60. N.F. Zhang, (with M.T. Postek, R.D. Larrabee, L. Carroll, W.J. Keery), A New Algorithm for the Measurement of Pitch in Metrology Instruments, *Proceedings of SPIE*, 1996, vol. 2725, pp. 147–158.

6.1.2. NIST Technical Reports

1. C. Croarkin (with J. R. Ehrstein), Standard Reference Materials: The Certification of 100 mm Diameter Silicon Resistiviaty SRMs 2541 through 2547 using Dual-Configuration Four-Point Probe Measurements, NIST SP 260-131, August 1997, 84p.
2. E.S. Lagergren, (with R.R Zarr, M.K Kumaran) NIST/NRC-Canada Interlaboratory Comparison of Guarded Hot Plate Measurements: 1993-1997. NIST IR 6087, 1997.
3. M.G. Vangel (with W.J. Rossiter, E. Embree, K.M. Kraft, and J.F. Seiler), Performance of Tape-Bonded Seams of EPDM Membranes: The Effect of Loading on Creep Resistance Under Peel Stress, *NIST Building Science Series 175*, 1996.
4. M.G. Vangel (with K.L. Stricklett), Electric Motor Efficiency Testing Under the New Part 431 of Chapter II of Title 10, Code of Federal Regulations: Enforcement Testing, *NIST Technical Note 1422*, 1996.
5. Mark G. Vangel and James J. Filliben (with Walter J. Rossiter and Kevin M. Kraft), Performance Tape-Bonded Seams of EPDM Membranes: Effect of Material and Application Factors on Peel Creep-Rupture Response, *NIST Building Science Series, 176*, 1997.
6. M.G. Vangel (with K.L. Stricklett), Analysis of Proposals for Compliance and Enforcement Testing Under the New Part 431; Title 10, Code of Federal Regulations, *NISTIR 6092*, 1997.
7. C.M. Wang, (with K.B. Rochford), Uncertainty in null polarimeter measurements. NIST IR 96–5055, 1996, 16p.

6.1.3. Book Reviews

1. D.L. Banks, review of *Operational Subjective Statistical Methods* by F. Lad, in *Journal of Classification*, 14(2), 1997, p. 78-79.
2. D.L. Banks, review of *Advances in Social Network Analysis* by S. Wasserman and J. Galaskiewicz, in *Journal of Classification*, 13(2), 1996, p. 356-358.

3. D.L. Banks, Strategies for superlarge datasets, *Newsletter of the Classification Society of North America*, #50, July 1997.
4. D.L. Banks, The geometry of location estimators, *Newsletter of the Classification Society of North America*, #49, April 1997.
5. D.L. Banks, The problem of excess genius, *Newsletter of the Classification Society of North America*, #48, February 1997.
6. M.G. Vangel, Statistical Reliability Theory: Models and Statistical Methods, A. Høyland and M. Rausand, *Technometrics*, 38(1), 1996, pp. 79-80.

6.1.4. Publications in Process

1. D.L. Banks (with S. Kotz and C.B. Read), *Encyclopedia of Statistical Sciences, Update Volume 3.*
2. D.L. Banks (with G. Constantine), Metric models for random graphs, to appear in *Journal of Classification.*
3. D.L. Banks (with G. Constantine, A. Merriwether, R. LaPlante), Analysis of pairwise differences of DNA sequences, to appear in *Journal of Nonparametrics.*
4. D.L. Banks (with R.J. Olszewski) Estimating local dimensionality, to appear in *Proceedings of the Statistical Computing Section of the American Statistical Association*, 1997.
5. D.L. Banks (with W. Shannon) A distance metric for classification trees, to appear in the *Proceedings of the 1997 Conference on Artificial Intelligence and Statistics.*
6. D.L. Banks (with W. Shannon) Combining classification trees using MLE, submitted to *Statistics in Medicine.*
7. K.J. Coakley, D.S. Simons, Detection and Quantification of Isotopic Inhomogeneity
8. K.J. Coakley, Statistical Planning for a Neutron Lifetime Experiment Using Magnetically Trapped Neutrons
9. J.J. Filliben, (with K.G.W. Inn, Z.C. Lin, M.K. Schultz, W.C. Burnett, G. Smith, S.R. Biegalski), Identification of radionuclide partitioning in soils and sediments: determination of best settings for the exchangeable fraction of the NIST standard sequential extraction protocol, *Applied Radiation and Isotopes*, submitted.
10. W.F. Guthrie (with M.W. Cresswell, N.M.P. Guillaume, R.A. Allen, R.N. Ghoshtagore, J.C. Owen III, Z. Osborne, N. Sullivan, and L.W. Linholm), Extraction of Sheet-Resistance from Four-Terminal Sheet Resistors Replicated in Monocrystalline Films having Non-Planar Geometries, *Proc. of IEEE International Conf. on Microelectronic Test Structures*, Kanazawa, Japan, March 23-26, 1998.
11. C. Hagwood (G. Mulholland, Y. Sivathanu). The DMA Transfer Function with Brownian Motion, a Trajectory/Monte-Carlo Approach, *Journal of Aerosol Science and Technology*, accepted.
12. C. Hagwood. A Probabilistic Approach to Exit from a Potential Well, *Stochastic Processes and their Applications*, submitted.
13. A. Heckert, (with E. Simiu and T. Whalen), Estimates of Hurricane Wind Speeds by the "Peaks Over Threshold" Approach, *Journal of Structural Engineering*, to appear.

14. A. Heckert, (with E. Simiu), Wind Direction and Hurricane-Induced Ultimate Wind Loads, *International Journal of Wind Engineering and Industrial Dynamics*, to appear.
15. A. Heckert, (with J.J.. Filliben), DATAPLOT Reference Manual, Volume I: Commands, *NIST Handbook 148*, to appear.
16. A. Heckert, (with J.J.. Filliben), DATAPLOT Reference Manual, Volume II: LET Subcommands and Library Functions, *NIST Handbook 148*, to appear.
17. E.S. Lagergren, (with M.J. Simon, K.A. Snyder) Concrete Mixture Optimization Using Statistical Mixture Design Methods, *Proc. of PCI/FHWA International Symposium on High-Performance Concrete*, accepted.
18. E.S. Lagergren, (with D.P. Bentz, E.J. Garboczi) Multi-Scale Microstructural Modelling of Concrete Diffusivity: Identification of Significant Variables, *Cement, Concrete, and Aggregates*, accepted.
19. E.S. Lagergren, (with R.R. Zarr) Development of Thermal Insulation Standard Reference Materials Using "Good" Experiment Design", *Proc. of 24th International Thermal Conductivity Conference and 12th International Thermal Expansion Symposium*, accepted.
20. E.S. Lagergren, (with P.Over) Comparing Interactive Information Retrieval Systems Across Sites: The TREC-6 Interactive Track Matrix Experiment,*21st International ACM SIGIR Conference on Research & Development in Information Retrieval*, submitted.
21. E.S. Lagergren, (with S.G. Malghan, R.S. Premachandran, R.K. Khanna) An Improved Method of Silicon Nitride Powder Processing, *Powder Technology*, submitted.
22. H.K. Liu, (with J.T. Hwang) High-dimensional empirical linear prediction with application to quality assurance in industrial manufacturing, submitted to *Technometrics.*
23. H.K. Liu, (with D.L. Duewer, S. Niezgoda, D.J. Reeder) Graphical tool for RFLP measurement quality assurance: single-locus tolerance charts, under submission.
24. M.G. Vangel, ANOVA Estimates of Variance Components for Quasi-Balanced Mixed Models, *Journal of Statistical Planning and Inference*, to appear.
25. M.G. Vangel and A.R. Rukhin, Maximum-Likelihood Analysis for Heteroscedastic One-Way Random Effects ANOVA in Interlaboratory Studies, *Biometrics*, to appear.
26. M.G. Vangel and A.L. Rukhin, Maximum-Likelihood Estimation for a One-Way Heteroscedastic ANOVA Model, With Applications to Interlaboratory Studies, *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, to appear.
27. M.G. Vangel, Saddlepoint Approximations for the Joint Distribution of the Mean and an Extremum, With Applications to Quality Control, *Proceedings of the Third Army Conference on Applied Statistics*, U.S. Army Research Laboratory, to appear.
28. Mark G. Vangel (with Walter J. Rossiter and Kevin M. Kraft), Performance of Tape-Bonded Seams of EPDM Membranes: Comparison of the Creep-Rupture Performance of Tape-Bonded and Adhesive-Bonded Samples as Affected by Test Temperature, Exposure Prior to Testing, Cold Temperature Preparation, and Shear

Loading, *NIST Building Science Series*, to appear.

29. Mark G. Vangel (with Walter J. Rossiter and Kevin M. Kraft), Performance of Tape-Bonded Seams of EPDM Membranes: a Field Investigation, *Proceedings of the Xth Congress of the International Waterproofing Association*, to appear.
30. M.G. Vangel, One-Sided β-Content Tolerance Limits for Mixed Models With Two Components of Variance, *Technometrics*, submitted.
31. A.L. Rukhin, B.J Biggerstaff and M.G. Vangel, Restricted Maximum Likelihood Estimation of a Common Mean and the Mandel-Paule Algorithm, *Journal of Statistical Planning and Inference*, submitted.
32. C.M. Wang, (with C.A. Hamilton), The fourth interlaboratory comparison of 10 V Josephson voltage standards in North America, *Metrologia*, 35, (1), 1998, in press.
33. C.M. Wang, (with J.R. Juroshek, G.P. McCabe), Statistical analysis of network analyzer measurements, CAL LAB, *The International Journal of Metrology*, to appear.
34. C.M. Wang, (with P.A. Williams), Corrections to fixed analyzer measurements of polarization mode dispersion, *Journal of Lightwave Technology*, to appear.
35. N.F. Zhang, Estimating process Capability Indices for Autocorrelated Data, *Journal of Applied Statistics*, to appear.
36. N.F. Zhang, Combining Process Capability Indices, *1997 Proceedings of the Section on Quality and Productivity of American Statistical Association*, to appear.
37. N.F. Zhang, A Multivariate Exponentailly Weighted Moving Average Control Chart for Stationary process Data, submitted for publication.

6.1.5. Working Papers

1. D.L. Banks, R.C. Hagwood, R. Kacker (with W. Dashiell, L. Gallagher and L. Rosenthal), Software testing by statistical methods.
2. D.L. Banks (with R. Maxion and R.J. Olszewski) Comparing methods for nonparametric regression. Under revision for *Journal of Computational and Graphical Statistics.*
3. D.L. Banks (with T. Kurfess and L.J. Wolfson) Assessing conformance to geometric tolerance. In preparation for *Technometrics.*
4. K.J. Coakley, D. Simons, G.Lamaze, H.Chen-Mayer, Comparison of SIMS and NDP measurements
5. K.J. Coakley, S. Dewey, Simulation of Neutron Scattering and Rocking Curve Data
6. J. Filliben, L. Gill, W. Guthrie, E. Lagergren, J. Rogers, M. Vangel; StRD: Statistical Reference Datasets for Assessing the Numerical Accuracy of Statistical Software, NIST Technical Note 1396, 1997.
7. L.M. Gill, (with K.E. Sharpless, S.A Margolis, and S.A. Wise) Certification of Nutrient Concentration in Standard Reference Material 2383: Baby Food Composite, *Journal of the Association of Official Analytical Chemists (AOAC) International* submitted
8. L.M. Gill (with K.E. Sharpless, M. Arce-Osuna, and J. Brown Thomas) Value

Assignment of Retinol, Retinyl Palmitate, Tocopherol, and Carotenoid Concentrations in Standard Reference Material 2383: Baby Food Composite, *Journal of the Association of Official Analytical Chemists (AOAC) International* submitted

9. W.F. Guthrie (with S. Mayo and J.J. Kopanski), Intermittent-Contact Scanning Capacitance Microscopy Imaging and Modeling for Overlay Metrology *Proc. of Int. Conf. on Characterization and Metrology for ULSI Technology*, Gaithersburg, MD, March 23-27, 1998.
10. H.K. Liu, (with J.T. Hwang) Testing for nonmodal error using high-dimensional empirical linear prediction.
11. M.G. Vangel, (with D.G.M. Anderson), Richardson's Algorithm and the Approximate Solution of Singular and Inconsistent Matrix Equations.
12. M.G. Vangel, The Joint Distribution of the Mean and an Extremum of a Normal Sample.
13. M.G. Vangel, A.L. Rukhin and S.D. Leigh (with G.B. McKenna, B. Lotz and C. Straupe) The τ-Effective Paradox Revisited: An Extended Analysis of Kovacs' Volume Recovery Data for Poly(Vinyl Acetate).
14. M.G. Vangel (with F.W. Scholz), Tolerance Bounds and C_{pk} Confidence Bounds Under Batch Effects, in *Statistics for Industry and Technology*, Birkhaeuser, New York, 1998.
15. Mark G. Vangel (with Dianne L. Poster, Maria J. Lopez De Alda, Michele M. Schantz, Lane C. Sander, and Stephen A. Wise), Certification of a Diesel Particulate Related Standard Reference Material (SRM 1975) For PAHs.
16. Mark G. Vangel (with Dianne L. Poster, Michele M. Schantz, and Stephen A. Wise), Analysis of Standard Reference Material 1649A, Organics in Urban Dust, for the Determination of Chlorinated Organic Contaminants Associated With Atmospheric Particulate Matter, submitted to *Environmental Science and Technology*
17. N.F. Zhang, Comparisons of Control Charts for Autocorrelated Data.

6.1.6. Acknowledgements in Publications

1. C. Croarkin in: Z. M. Zhang, T. R. Gentile, A. L. Migdall, R. U. Datla, Standard Reference Materials:Transmission Filters with Measured Optical Density at 1064 nm Wavelength - SRMs 2046, 2047, 2048, 2049, 2050, 2051, NIST SP 260-128, 1997.
2. W.F. Guthrie acknowledged in: B.W. Mangum, E.R. Pfeiffer, G.F. Strouse, J. Valencia-Rodriguez, J.H. Lin, T.I. Yeh, P. Marcarino, R. Dematteis, Y. Liu, Q. Zhao, A.T. Ince, F. Cakrioglu, H.G. Nubbemeyer, H.-J. Jung, Comparisons of Some NIST Fixed-Point Cells with Similar Cells of Other Standards Laboratories, *Metrologia*, 1996, 33, pp. 215-225.
3. M.S. Levenson acknowledged in: S.M. Stigler, Statistics and the Question of Standards, *Journal of Research of the National Institute of Standards and Technology*, Vol 101, 1996, pp. 779-789.

6.2. Talks

6.2.1. Technical Talks

1. D.L. Banks, Four approaches to TREC data, TREC Conference, NIST, November 1997.
2. D.L. Banks, NAIVE: A postprocessor for network anomaly/intrusion detectors, Symposium on Visualization and Testing in Intrusion Detection, NIST, November, 1997.
3. D.L. Banks, Strategies for superlarge datasets, NIST Seminar, November, 1997.
4. D.L. Banks, Issues in datamining, University of Waterloo IIQP Colloquium Series, October, 1997.
5. D.L. Banks, Mining superlarge datasets, SRI Intrusion Detection Workshop, July 1997.
6. D.L. Banks, Analysis of superlarge datasets, annual meeting of the Classification Society of North America, June 1997.
7. D.L. Banks, Preanalysis of superlarge superlarge datasets, DIMACS Workshop, Rutgers University, May 1997.
8. D.L. Banks, Problems in random number generation, DARPA Intrusion Detection Workshop, February 1997.
9. K.J Coakley, Statistical Planning for Neutron Lifetime Experiment, Technical Collaboration Meeting at NIST, July 1997
10. K.J Coakley, Optimal Design of Neutron Lifetime Experiment, Joint Meetings of American Statistical Society, Anaheim, August 1997
11. K.J Coakley, Aerosol Concentration in a New Spectrometer, K.J.Coakley and K.Ehara, American Aerosol Association, Denver, October, 1997.
12. K.J Coakley, Optimal Design of Neutron Lifetime Experiment, Technical Collaboration Meeting at Harvard University, October, 1996
13. K.J Coakley, Chaotic Behavior of Marginally Trapped Neutrons, Technical Collaboration Meeting at Harvard University, October, 1996
14. C. Croarkin, ISO and NIST Uncertainty Policies, Laser Measurements Workshop, Boulder, CO, Aug. 15, 1997.
15. J.J. Filliben, (with K.G.W. Inn and Z. Lin), Certification of the NIST Ocean Sediment SRM Using Non-normal Parametric Statistic Analysis, Physics Laboratory Ionizing Radiation Division Seminar, Gaithersburg, MD, February 25, 1997.
16. J.J. Filliben, Evaluation of Uncertainty in Scientific Measurement, Advanced School of Metrology in Brazil: Evaluation of Uncertainty in Measurement, Angra dos Reis, Brazil, March 3, 1997.
17. J.J. Filliben and L.M. Gill, Summary Analysis: Effect of Grinding Parameters on SRBSN, RBSN, and SSN, NIST Ceramic Machining Consortium, 10th Program Review Meeting, Washington, D.C., April 10, 1997.
18. J.J. Filliben, C. Croarkin, W.G. Guthrie, A. Heckert, (with P.A. Tobias, P.D. Spagon, B. Hembree, J. Prins, and C. Zey), NIST/SEMATECH Engineering Statistics Handbook Demo, 14th Annual Statistical Methods Symposium, San Antonio, TX,

May 6, 1997.

19. J.J. Filliben, NIST/SEMATECH Engineering Statistics Handbook: Dataplot Highlights, 14th Annual Statistical Methods Symposium SEMATECH Statistical Methods Advisory Council, San Antonio, TX, May 7, 1997.
20. J.J. Filliben, Examples for Teaching Design of Experiments, 1997 Quality and Productivity Research Conference: Design of Experiments and Industrial Statistics, Orlando, FL, May 11, 1997.
21. J.J. Filliben, W.G. Guthrie, A. Heckert, Integration of Statistical Optimization Tools in Process/Product Design, SIMA / HPCC Panel, Gaithersburg, MD, July 18, 1997.
22. J.J. Filliben, A. Heckert, M.G. Vangel, Material Basis Values Within a Complete Statistics Package: RECIPE (REgression Confidence Intervals for PErcentiles) in DATAPLOT. Composite Materials Handbook (MIL-HDBK-17) Coordination Meeting, Williamsburg, VA, September 10, 1997.
23. J.J. Filliben, C. Croarkin, W.G. Guthrie, A. Heckert, M. Reeder, J. Wang, (with P.A. Tobias, P.D. Spagon, B. Hembree, J. Prins, and C. Zey), NIST/SEMATECH On-Line Statistics Handbook for Engineers and Scientists, American Statistical Association, Anaheim, CA, August 10, 1997.
24. J.J. Filliben, (with K.C. Li), A Systematic Approach to the Analysis of Complex Interaction Patterns in a 2-Level Factorial Design, Fall Technical Conference, Baltimore, MD, October 17, 1997.
25. J.J. Filliben, and A. Heckert, An Overview of the Dataplot Graphics and EDA Software System, Washington Statistical Society, Washington, D.C., December 10, 1997.
26. J.J. Filliben, Graphical Analysis of 2-Factor Data, Part of the 5-Day Course: Analysis of Variance, by S.D. Leigh, Statistics for Scientists Seminar Series, Gaithersburg, MD, January 21, 1998.
27. J.J. Filliben, Graphical Analysis of 3-Factor Data, Part of the 5-Day Course: Analysis of Variance, by S.D. Leigh, Statistics for Scientists Seminar Series, Gaithersburg, MD, February 4, 1998.
28. J.J. Filliben, Graphical Analysis of 4-Factor Data, Part of the 5-Day Course: Analysis of Variance, by S.D. Leigh, Statistics for Scientists Seminar Series, Gaithersburg, MD, February 11, 1998.
29. L.M. Gill, Navigating Through the Statistical Reference Dataset (StRD) Website, Shaping Statistics for Success in the 21st Century, Joint Statistical Meetings '97, Anaheim, CA, August 11, 1997.
30. L.M. Gill, Summary Analysis: High Performance Ceramic Experiment to Characterize the Effect of Grinding Parameters on Sintered Reaction Bonded Silicon Nitride, Reaction Bonded Silicon Nitride, and Sintered Silicon Nitride, NIST Ceramic Machining Consortium, 10th Program Review Meeting, Washington, D.C., April 10, 1997.
31. W.F. Guthrie, Selection and Certification of Reference Datasets for Testing the Numerical Accuracy of Statistical Software for Linear Procedures, NIST, Boulder, CO, September 26, 1997.
32. W.F. Guthrie, NIST/SEMATECH Online Statistics Handbook for Engineers and

Scientists, Joint Statistical Meetings, Anaheim, CA, August 10, 1997.

33. W.F. Guthrie, Selection and Certification of Reference Datasets for Testing the Numerical Accuracy of Statistical Software for Linear Procedures, Joint Statistical Meetings, Anaheim, CA, August 11, 1997.
34. W.F. Guthrie, Selection and Certification of Reference Datasets for Testing the Numerical Accuracy of Statistical Software for Linear Procedures, NIST, Gaithersburg, MD, July 24, 1997.
35. W.F. Guthrie, Demonstration of the NIST/SEMATECH Engineering Statistics Handbook, SEMATECH Statistical Methods Symposium, San Antonio, TX, May 6, 1997.
36. A. Heckert, (with E. Simiu), Wind Direction and Hurricane-Induced Ultimate Wind Loads, *Proceedings of the 2ND European and African Conference on Wind Engineering*, Genova, Italy, 6/97.
37. E.S. Lagergren, A Comparison of Approaches for Robust Parameter Design, Third U.S. Army Conference on Applied Statistics, George Mason University, Fairfax, VA, October 24, 1997.
38. E.S. Lagergren, Statistical Reference Datasets (StRD) - An Overview, NIST, Boulder, CO, September, 26, 1997.
39. E.S. Lagergren, Statistical Reference Datasets (StRD) - An Overview, Joint Statistical Meetings, Anaheim, CA, August 11, 1997.
40. E.S. Lagergren, Statistical Reference Datasets (StRD) - An Overview, NIST, Gaithersburg, MD, July 24, 1997.
41. M.S. Levenson, Spectral Methods for Data Analysis, NIST, Gaithersburg, MD, June, 1997.
42. M.S. Levenson, Spectral Methods for Data Analysis, NIST, Boulder, CO, November, 1997.
43. M.S. Levenson, Toolbox of Techniques for Uncertainty Analysis: Uncertainty in Linear Calibration, Measurement Science Conference, Pasadena, CA, February, 1998.
44. M.S. Levenson, Case Studies from NIST: Radiometric Measurements, Measurement Science Conference, Pasadena, CA, February, 1998.
45. M.S. Levenson, Statistical Considerations in Calibration, Baltimore Section of American Society for Quality, Baltimore, MD, February, 1998.
46. H.K. Liu, High-dimensional empirical linear prediction, the 1997 Joint Statistical Meetings, August 12, 1996, Anaheim, CA.
47. H.K. Liu, Quality assurance in DNA fingerprinting, Fall Meeting of the Louisiana Chapter of the American Statistical Association, October 24, 1997, New Orleans, LA.
48. M.G. Vangel, Maximum-Likelihood Estimation for a One-Way Heteroscedastic ANOVA Model, With Applications to Interlaboratory Studies, Joint Statistical Meetings, Anaheim, CA, August, 1997.
49. M.G. Vangel, Saddlepoint Approximations for the Joint Distribution of the Mean and an Extremum, With Applications to Quality Control, Army Conference on Applied Statistics, Fairfax, VA, October, 1997.
50. M.G. Vangel, Combining Information in Interlaboratory Studies, Boston University,

February, 1998.

51. C.M. Wang, Fourth interlaboratory comparison of 10 V Josephson voltage standards, National Conferences of Standards Laboratories Workshop and Symposium, Atlanta, GA, July 27–31, 1997.
52. C.M. Wang, Robust regression applied to optical fiber dimensional quality control, Joint Statistical Meetings, Anaheim, CA, August 10–14, 1997.
53. N.F. Zhang, Combining Process Capability Indices, Joint Statistical Meetings, Los Angles, CA, August 12, 1997
54. N.F. Zhang, Statistical Quality Control and Its Applications in Portfolio Risk Management, Department of Statisitcs, Nanjing Institute of Economics, Nanjing, China, June 3, 1997.
55. N.F. Zhang, Statistical Control Charts for Stationary Process Data, Department of Mathematical Statistics, East China Normal University, May 28, 1997.
56. N.F. Zhang, A Multivariate Control Chart for Stationary Process data, International Symposium on Contemporary Multivariate Analysis and Its Applications, Hong Kong, May 20, 1997.
57. N.F. Zhang, A Statistical Measure for the Sharpness of SEM Images, SPIE's 22nd International Symposium on Microlithography, Santa Clara, CA, March 11, 1997.

6.2.2. General Interest Talks

1. K.J Coakley, Nonequilibrium Kinetics of Neutral Atoms in a Harmonic Potential, University of Chile, Physics Department, March, 1996
2. K.J Coakley, A Bootstrap Method for Nonlinear Filtering of EM-ML Reconstructions of PET Images, Catholic University, Santiago, Chile, Mathematics Department and University of Chile, Applied Mathematics Department, March, 1996
3. J.J. Filliben, Basketballs, Funnels, and Designed Experiments, Adventures in Science, NIST, Gaithersburg, MD, January 24, 1998.
4. L.M. Gill, Navigating Through the Statistical Reference Dataset (StRD) Website, Statistical Engineering Division Seminar Series, NIST, Gaithersburg, MD, July 24, 1997.
5. L.M. Gill, Navigating Through the Statistical Reference Dataset (StRD) Website, Mathematical and Computational Sciences Division Seminar Series, NIST, Boulder, CO, September 26, 1997.

6.2.3. Workshops for Industry

1. C. Croarkin (with G. Harris), Advanced Workshop on Mass Measurements, NIST, Gaithersburg, Mar. 10-12, 1997.
2. C. Croarkin (with J. Keller), Seminar on Mass Intercomparisons consisting of a series of lectures on statistical procedures for mass calibrations for metrologists from the CAMET region, sponsored by the Organization of American States and the Ministry of Commerce and Industries in Panama, Panama City, Panama, May. 21-23, 1997.

3. C. Croarkin, M. Levenson, C. M. Wang, Workshop on Easy and Not-So-Easy Methods of Uncertainty Analysis given in conjunction with the 1998 Measurement Science Conference, Doubletree Hotel, Pasadena, California, Feb. 2-3, 1998.
4. J.J Filliben, L.M. Gill, E.S. Lagergren, Workshop on Improving Product and Process Quality Using Experiment Design, NIST, Gaithersburg, MD, December 1-5, 1997.

6.2.4. Lecture Series

1. D.L. Banks, Shortcourse on nonparametric regression, annual meeting of the Classification Society of North America, June 1997.
2. J.J. Filliben, Exploratory Data Analysis, Statistics for Scientists Seminar Series, Gaithersburg, MD, November 14/18/25 and December 9/16, 1997.
3. M.G. Vangel, *Tolerance Intervals and Tolerance Regions*, tutorial for the Construction Materials Research Laboratory, NIST, May 17, 1997.

6.3. Professional Society Activities

6.3.1. NIST Committee Activities

1. K.J Coakley, served on NIST Research Advisory Committee
2. K.J Coakley, served on Boulder Editorial Review Board
3. C. Croarkin, Member, NIST Standards Advisory Committee.
4. C. Croarkin, Member, EEEL CALCOM Committee on Resistivity.
5. L.M. Gill, Statistical Consultant, NIST Ceramic Machining Consortium.
6. C.M. Wang, Member, EEEL MCOM Committee on Calibration Service for laser power and energy measurements at the laser wavelength of 248 nm.
7. C.M. Wang, Member, EEEL MCOM Committee on wavelength reference absorption cell SRM.

6.3.2. Standards Committee Memberships

1. K.J Coakley, served on Calcom committee for SRM 2538(Coplaner Waveguide)
2. C. Croarkin, Vice-Chair, US TAG to ISO TC-69 on Statistical Methods.
3. C. Croarkin, Convenor of WG7 of ISO/TC69/SC6.
4. C. Croarkin, Chair, ANSI ASC Statistics Subcommittee.
5. C. Croarkin, Member, ASTM E-11 Subcommittee on Quality and Statistics.

6.3.3. Other Professional Society Activities

1. D.L. Banks, Secretary, International Federation of Classification Societies, 1997-2001.

2. D.L. Banks, Board member, Classification Society of North America, 1996-1998.
3. D.L. Banks, Session chair for Intrusion Detection and Computer Security, annual meeting of the Classification Society of North America.
4. D.L. Banks, Member of the American Statistical Society Committee on Scientific Freedom and Human Rights, 1997-2000.
5. D.L. Banks, Session chair for Design and Estimation for Establishment Surveys, joint statistical meetings.
6. D.L. Banks, Session chair for Application of Mixed Models in Calibration, joint statistical meetings.
7. D.L. Banks, Editorial board, *CSNA Service*, 1996-2000.
8. J. Rosenblatt, Chairman of the Youden Award Committee of the American Statistical Association.
9. L.M. Gill, Secretary / Treasurer of the ASA Quality and Productivity Research Conference Steering Committee, 1996-1999.
10. E.S. Lagergren, Steering Committee Member, ASA Quality and Productivity Research Conference, 1995–1997.
11. E.S. Lagergren, Program Chair, ASA Quality and Productivity Section, 1997–1998.
12. M.G. Vangel, Program-chair elect, Section on Risk Analysis, American Statistical Association.
13. M.G. Vangel, Awards committee chairman, Section on Physical and Engineering Sciences, American Statistical Association.

6.4. Honors

1. W.F. Guthrie, Department of Commerce Bronze Medal, December 1997.
2. R.N. Kacker, The S-Check software performance improvement tool was selected for a 1997 R&D 100 Award as one of the year's 100 most technologically significant products. Joint work with G. Lyon, R. Snelick, J. Antonishek, M. Courson, N. Drouin, M. Indovina, J. Ja'Ja' and D. Rodriguez.
3. E.S. Lagergren, Standard Reference Data Program Award, November 1997.

6.5. Trips Sponsored by Others and Site Visits

1. D.L. Banks, Visit to University of Waterloo IIQP Center and Department of Statistics, October, 1997.
2. D.L. Banks, Visit to SRI, August, July, 1997.
3. D.L. Banks, Visit to the Center for Discrete Mathematics, Rutgers University, May, 1997.
4. M.G. Vangel and J.J. Filliben, Visit to Composite Materials Handbook (MIL-HDBK-17) Coordination Meeting, Williambsburg, VA, September 10, 1997.
5. H.K. Liu, Fall Meeting of the Louisiana Chapter of the American Statistical Association, October 24, 1997, New Orleans, LA.
6. M.G. Vangel and J.J Filliben, *Composite Materials Handbook Coordination Group*

Meeting, Williamsburg, VA, Oct. 1997. Trip sponsored by the U.S. Army Research Laboratory.

7. N.F. Zhang, Invited visit to East China Normal University in Shanghai, China, where a workshop titled "Applied Time Series Analysis" and a talk titled "Statistical Process Control Charts for Autocorrelated Data" were given, May 26–30, 1997.
8. N.F. Zhang, Invited visit to Nanjing Institute of Economics in Nanjing, China, where a talk titled "Statistical Quality Control and Its Applications in Portfolio Risk Mangement" was given, June 2–5, 1997.

6.6. Professional Journals

6.6.1. Editorships

1. M.G. Vangel, Associate Editor, *Technometrics.*

6.6.2. Refereeing

1. D.L.Banks, *Journal of Mathematical Psychology*, *Journal of Mathematical Sociology*, *Psychometrika*, *Sociological Methodology*, *Statistics in Medicine*, *Technometrics.*
2. K.J. Coakley, IEEE Transactions on Medical Imaging, Journal of the American Statistical Society, Computational Statistics and Data Analysis
3. R.C. Hagwood, Institute of Mathematics, Academia Sinica.
4. E.S. Lagergren, Journal of Quality Technology.
5. M.S. Levenson, *IEEE Transactions on Medical Imaging.*
6. M.S. Levenson, *INFORMS Journal of Computing.*
7. H.K. Liu, Annals of Statistics, Academia Sinica.
8. M.G. Vangel, Technometrics, Journal of the American Statistical Association, Statistica Sinica, Mathematical Methods of Statistics.
9. C.M. Wang, Communications in Statistics.
10. N.F. Zhang, Computational Statistics and Data Analysis.

6.7. Proposal Reviewing

1. D.L. Banks, Agency for International Development, National Science Foundation.
2. C. Hagwood, Fulbright Fellowship

6.8. Training & Educational Self-Development

1. D.L. Banks, John Musa's shortcourse on Software Reliability, June 24, 1997.

2. D.L. Banks, Bruce Shneier's course on Cryptography, September 29-30, 1997.
3. L.M. Gill, Data Analysis Using S-Plus, Washington, DC, Sept. 22-25, 1997.
4. W.F. Guthrie, Advanced Design of Experiments, SEMATECH, May 8-9, 1997.
5. E.S. Lagergren, Data Analysis in S-Plus, Washington, DC, September 22-25, 1997.

6.9. Special Assignments

1. E.S. Lagergren, M.C. Croarkin, J.J. Filliben, L.M. Gill, W.F. Guthrie, H.K. Liu, M.G. Vangel, N.F. Zhang (with P. Fagan, J.E. Rogers, B.W. Rust), Statistical Reference Datasets Team, 1997.
2. E.S. Lagergren, SRM Team Leader, 1997.
3. M.G. Vangel, Chairman, Statistics Working Group, Mil-Handbook-17 (Composite Materials Handbook) Coordination Group.
4. M.G. Vangel, Program Committee Member, Army Conference on Applied Statistics.