Journal of Official Statistics, Vol. 26, No. 3, 2010, pp. 443–453

Statistical Careers in United States Government Science Agencies

Nell Sedransk[1]

The role of statistics in those U.S. government agencies that focus on progress in science and engineering became prominent at the end of the Second World War. The success of statistics in that historical period came from the power of statistics to enable science to advance more rapidly and with great assurance in the interpretation of experimental results. Over the past three quarters of a century, technology has changed both the practice of science and the practice of statistics. However, the comparative advantage of statistics still rests in the ability to achieve greater precision with fewer errors and a deeper understanding. Examples illustrate some of the challenges that complex science now presents to statisticians, demanding both creativity and technical skills.

Key words: History of statistics; complex system models; engineering statistics; metrology; high-dimensional data.

1. Introduction

The clear mission of statistics in government science agencies is, and has always been, to enable scientists to make progress and to accelerate scientific advances through development and implementation of statistical methodology.

In the United States, the science agencies are generally separate from the statistical agencies that are responsible for regular reporting of national statistics, although a few agencies have separate science and statistical branches. Thus for example, the National Institutes of Health (NIH) and the National Cancer Institute (NCI), the largest complex of science agencies, are separate from the National Center for Health Statistics (NCHS), part of Center for Disease Control (CDC). The *nonmedical* science agencies are primarily oriented toward the physical science and engineering with foci on geology, space, weather, climate and atmospheric science, engineering, metrology, and the basic sciences.

The importance of this statistical role in the *nonmedical* science agencies emerged as part of the war efforts for the Second World War. Since then, this role has taken a wide variety of forms. On the intellectual side, there has been a continually modulating balance between statistical research and direct assistance to scientific applications. On the logistical side, the position of statisticians has cyclically alternated between grouping statisticians together and embedding them individually in research teams. Dramatic changes in the scientific research paradigm over twenty-year periods have brought about equally important changes in the specific contributions of Statistics. At each point in time,

[1] National Institute of Statistical Sciences, P.O. Box 14006, Research Triangle Park, NC 27709-4006, U.S.A. Email: sedransk@niss.org

the driving forces for statistical applied research and application have reflected the research paradigm of the time so that the dominant statistical developments could respond to the scientists' perceived needs. In consequence the roles for statisticians have morphed from technical support to co-investigator. Amazingly perhaps, the mission today *is* the original mission; and over the years the challenge of that mission has only become greater – and more exciting.

Beginning during the Second World War, the government's physical science agencies became acutely aware of the need for efficiency in gathering and understanding information and of the need to deal with the uncertainty arising from imperfect measurement as well as random variation. Statistics offered disciplined and reproducible methods for understanding data from physical systems, and methodologies developed for scientists to implement while the scientists focused their creative energies on their sciences.

Today, it is almost as if the revolutions in science have brought the role of statistics full-circle, so that the opportunities are once again to develop new statistical theories and methodologies that will enable scientists to see beyond what their own competence can illuminate, beyond reports that easily accessible statistical software generates.

2. Historical Perspective – Half a Century

The Second World War had brought together statisticians, including now-legendary figures like John Tukey and Fred Mosteller, with pressing real applications and their urgent needs for analysis and information. By the 1950s, the idea that statistical theory and methods were a key to scientific efficiency had taken root in the science agencies with missions in the physical sciences and engineering. The driving force for the burst of statistical activities during this era came from the exploitation of statistical testing, modeling and prediction during and following the war. Notions of uncertainty needed to be codified to describe measurements and to identify primary sources of heterogeneous quality. The desire to transfer the efficiencies gained in agricultural experimentation to engineering and science applications focused attention on experimental design and linear models, especially analysis of variance components.

So the science agencies became hubs for the development of both the theory and methods that could be applied to engineering projects and to physical science experiments. Prominent statisticians from all over came to the science agencies to be part of these developments; they came as high school interns, as students, as research faculty from eminent universities, and some came to stay as permanent research statisticians and collaborators (Statistical Engineering Division, personal communication). NIH was a similar magnet for statisticians focused on medical research and biostatistics. Defense agencies and the research arms of the military supported their own research staffs and visitors and also made funding available to academic researchers both for fundamental research and for implementation.

It is difficult to explain to a generation that has never seen an electric calculator with or without automatic division capability, that statisticians were essential to scientists and engineers because the statisticians could use Abelian group theory to construct designs for individual experiments. Statisticians also focused a great deal of attention on formulas and mathematically clever approximations that would be computed manually, often using only paper and pencil. (A matrix representation for linear models theory was only being developed at this time.)

Statistics for experiments dominated the research on experimental design construction, leading to catalogues of factorial and fractional factorial designs and the first response surface designs for scientists to use. Analysis focused on linear models, components of variance, testing and rules for multiple tests, and sequential methods aimed at increasing efficiency. Approximations and calculation shortcuts were assembled into Mary Natrella's *Experimental Statistics, NBS Handbook 91* (Natrella 1963), designed as a compendium and resource for statisticians and scientists alike.

Statisticians in the physical science agencies, from earth and planetary science to engineering and metrology, included both researchers and computation staff. Generally, scientists' expertise did not extend either to designing an experiment or to the statistical analysis of the resulting data. So statisticians were collaborators, first to understand the science sufficiently to construct a suitable experimental design, and then to apply analytic methodology to the scientists' raw data and deliver both the statistical analyses and their interpretations.

By the 1970s computers and data bases were the driving force, not just for statistics but also for science. Statisticians were critical because data sets were large and difficult to navigate and statistical software was decidedly user-unfriendly. Exploratory Data Analysis offered the attractive possibility of exploring large data sets and modeling from *un*designed data bases. So the exciting areas of statistics revolved around data analysis and summarization, bringing new tools that for the most part required computers and computational expertise. This gave the statistician the *opportunity* to understand the science. Pattern detection techniques and EDA tools required statisticians' expertise for these data-driven analyses; so the scientist and the statistician became collaborators in developing conjectures.

Once statistical analysis had become the primary acceptable (if not required) form for presentation of scientific results, the statistical production staff of the 1970s and 1980s had to expand dramatically to meet the demand for generating analyses and reports requested by scientists and engineers. In a parallel development, statistical methodological research advanced. Newer statistical software was designed to be much more user-friendly, capable of analyzing more complex data and amenable to customizing in order to give scientists statistical tools to apply without the personal interface of a statistician.

In the 1990s the transition to computer-based science was realized. Pervasive computing was embodied in computerized equipment and automated data-recording of experimental data as well as preprocessing of raw data, so that the complexity of the formulae and algorithms became unimportant to the scientist-user. Statistical software for "standard statistical methodology" that had been the research goal a quarter of a century earlier was abundant. The new generation of scientists and engineers were completely at home with computer technology and computation; and computer scientists were rapidly developing alternative methods for exploration of data bases – "data mining" was becoming the rage. Algorithms still needed to be customized and macros needed to be written for special cases, but the statistician was no longer essential to construct an experimental design for a univariate or even a multi-variate outcome. Production analysis was being done by postdoctoral scientists and technicians; so that statisticians were no longer the interface between the laboratory and analyzed results.

What began in the 1990s to challenge and perplex the scientists and engineers was the extraction of information and synthesis of interpretation from high-dimensional data with complex internal structures that were not easily detectable through standard analyses or projections onto small subspaces. Interdisciplinary research and high-dimensional data dominated, requiring advances in computational statistics that were being made within and outside the science agencies. Development of new statistical methodology focused on computationally intensive statistics (e.g., Markov chain Monte Carlo), high-dimensional graphics and analysis, use of empirical probabilities computed from data bases, and Bayesian modeling and inference made possible with the new computing power and with computable and empirical distribution functions as well as classical closed form functions.

To fulfill the role of interdisciplinary collaborator, e.g., to embed stochastic elements into the scientific models, the statistician was now *required* to understand the science. As (deterministic) modeling from physical laws and/or curve fitting became more and more a primary tool for scientists and engineers, statisticians and applied probabilists held the critical expertise to integrate stochastic elements, especially into simulations used either to explore or to validate scientific conjectures.

3. Contemporary Perspective

The contemporary world, including the science agencies, is driven by BIG Science. Thus the same need for a disciplined approach to designing experiments and maximizing information from their analysis that drove statistical advances and statisticians' involvement in science following the Second World War has reemerged since 2000. Now it is scaled up by orders of magnitude.

BIG science is big in every way: multiple critical disciplines; massive, high-dimensional and often heterogeneous data; complex systems requiring complex and expensive experiments and equally complicated models.

For these ambitious scientific inquiries, the statistical questions that arise are the canonical questions that were also asked in the 1950s: how to design efficient experiments, how to analyze complex systems, understand complex interrelationships among contributing factors, determine the components of variation and make predictions, how to interpret and report high-dimensional results with statistically sound statements of their uncertainty. The dominant ideas are the understanding, modeling and inference for complex systems, uncertainty for system models, and representations (numerical or graphical) of high-dimensional or dynamic systems and processes and of the associated uncertainties.

Now that computational capabilities are not solely the purview of statisticians, complex (deterministic) system models are formulated and implemented by physical scientists and massive databases are mined by computer scientists. The question, "Why maintain a high-priced group of statisticians to 'do statistics'?" lurks in science agencies as they cope with real or *de facto* budget cuts. The answer has to come from the inclusion of stochastic components in models, stochastic bases for analyses, and probability-based statements of uncertainty. Experimental design is once again a pressing statistical area for research and the development of methodology. For research on complex systems and/or high-dimensional data, experimental design relies on simulation as well as on classical designs. Often experiments are interlinked so that the design paradigm comprises multiple serial or simultaneous experiments. Computation and computationally intensive statistics are pervasive – from experimental design construction to data mining to numerical solutions for non-linear model fitting.

Because the complex system models are often partially derived from physical laws and partially from empirical function-fitting, randomness does not often enter the model in a simple way. Rarely is the randomness of such a system captured in an additive error term; rarely are independence assumptions about distributions for the model parameters justified. Consequently statistical inferences, predictions and uncertainty may only be possible through computationally intensive methods; and for "hybrid models" that incorporate physical laws, black box, and statistical model components, the evaluation of model robustness is a necessary, yet difficult, part of each analysis.

And the role of the statistician is that of *co-scientist*; understanding the science is no longer optional.

The kinds of problems that statisticians now solve continue to include straightforward problems in a variety of areas, but the balance of their work is shifting toward specific scientific areas and specific engineering processes. Several examples follow. These are all drawn from Statistical Engineering Division (SED) projects at the National Institute of Standards and Technology (NIST), with descriptions to be found in SED (2003), SED (2004) and SED (2005).

4. Modeling Scientific Processes

Modeling complex equipment can require a series of experiments over several years to model the output and to characterize and compensate for both systematic and random errors (Wang et al. 2002; Clement et al. 2006; Hale et al. 2006; Williams et al. 2006). In the case of high-speed sampling oscilloscopes (critical instruments in the design of high-performance systems) the outputs are three wave-forms (the signal of interest plus two ancillary sinusoids); the jitter contributes random error while time-based distortion is systematic. Thus the solution draws broadly on statistical methodology from weighted nonlinear least squares to B-splines to parametric bootstrap. In Figure 1 the signal is shown as the scatter, the initially fitted (jitter-corrected) function as the wide line and the final function (corrected for time-based distortion) as the thin line.

Stochastic modeling also serves as the basis for designing experiments and for configuring equipment. For some nanotechnology experiments, the modeling of atomic and subatomic particle behavior is critical since the actual construction of the experimental venue may require very large expenditures both of money and of time, so that optimizing the experiment is critical both in terms of expected information and in terms of precision of the output. For example, the ability to control individual atoms ("atom-on-demand" technology) is the enabler for control over dopants in materials, for quantum communication and for the prospect of atom-by-atom construction. Modeling the number of atoms in a magneto-optical trap as a function of the loading rate and the loss

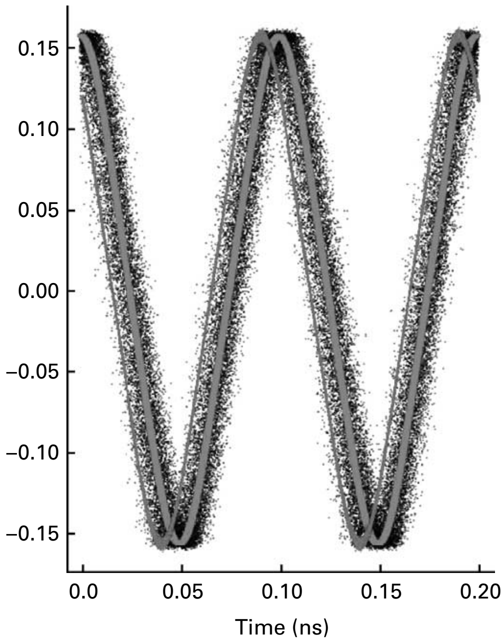

Fig. 1. Compensation for Random and Systematic Time-Base Distortion (courtesy of C.M. Wang, Statistical Engineering, P.D. Hale, Optoelectronics and D.F. Williams & K.A. Remley, Electromagnetics, NIST)

processes can be used to maximize the proportion of time a single atom spends in the trap. The goal is to determine the specifications for a prototype magneto-optical trap that optimizes the process; the statistical approach is to develop a theoretical model from which to simulate followed by simulation followed by experimental data. Assuming Poisson loading (parameter, λ) and mean time in trap before loss (parameter $1/\mu$), and making use of the Borel-Tanner distribution, the probability of exactly one atom in the trap as a function of λ and μ is shown in Figure 2 (SED 2003).

5. Understanding Complex Systems

At a higher, very ambitious level statistical reasoning as well as statistical methodology contributes to the process of exploring complex systems. Understanding the sequence of events in the collapse of the World Trade Center buildings is a case in point (Filliben 2005; NIST 2005). Statistical contributions of two different kinds proved crucial. The more straightforward was the use of experimental design principles to ensure that the testing of material properties using the precious few available samples of World Trade Center beams and structural materials yielded the most important information with the greatest precision possible. Experimental design was equally crucial to undertaking the final simulations of the collapse process by linking together four massive simulation packages (finite element analysis models plus a fire dynamics simulator) into a four-stage data chain. Since the size of simulation (including hundreds of thousands of finite elements) precluded a sufficient number of runs to examine all possible scenarios, classical statistical design principles proved essential.

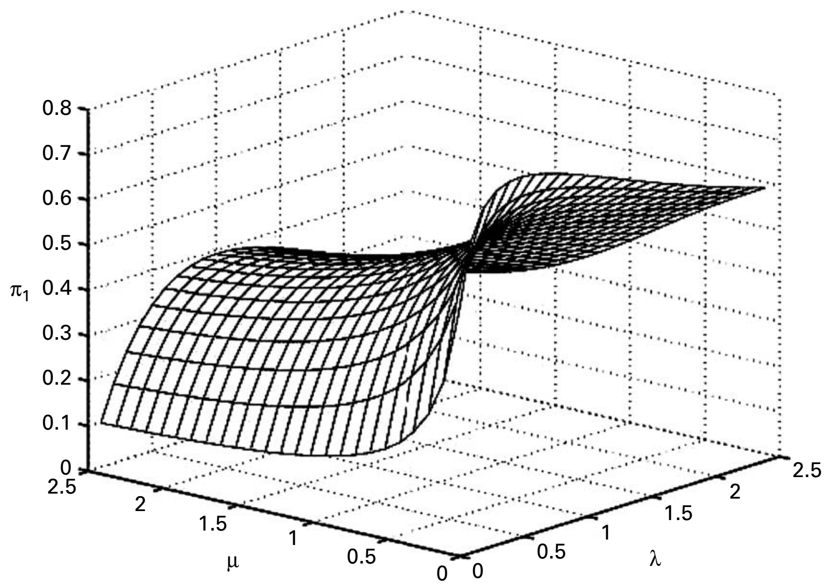

Fig. 2. Stochastic Model for the Number of Atoms in a Magneto-Optical Trap (courtesy of Andrew Rukhin, Statistical Engineering, NIST and Ionet Bebu, UMBC)

The second and greater contribution lay in using statistical principles to characterize the entire project in terms of the different kinds of information available: (i) data, (ii) simulated information (based on data-derived input parameters), (iii) expert information, and (iv) complex models tweaked to match the visible progression of the event. Primary data came in the form of video records from various perspectives (converted to second by second status of each window from initially intact to smoke then to flame and eventually to black before the collapse which was recorded by floor by stage using an even finer time grid). Auxiliary data came from architectural drawings, builder's specifications, and suppliers' lists of materials, down to specifications for individual beams at known locations in the structure. Experimental data came from laboratory tests on samples of structural materials from the collapsed buildings that were conducted at the National Institute of Standards and Technology. Models of complex failure modes came from several sources including fire models that had been developed over a decade or so and validated through their application to experimental fires and investigation of accidental fires. By creating an overview of all this input information, statistical principles were the key to understanding what was "knowable" from this extraordinary scientific investigation. These principles were the guideposts to distinguishing the conclusions that could be drawn with associated probability statements, the propositions that could be determined to be plausible or not based on the success of model recreation of the event, and the experts' judgments as well as assumptions that could only be examined in terms of robustness of conclusions. By understanding these boundaries on the conclusions, a highly efficient strategy for the entire project effort was possible – and was achieved.

6. Measuring and Quantifying Uncertainty

When the inference is about a single or vector parameter or prediction, uncertainty is represented by estimating a function of a second moment; and usually confidence intervals or highest posterior probability regions are quoted. However, for more complicated objects, attributes and processes the inference is more difficult, variation is more complicated and the quantification of uncertainty is a challenge. One such challenge is to measure and to quantify the perfection of a precision-made sphere, i.e., the discrepancy compared to an ideal, perfect sphere (Sedransk 2007). The difficulty is not the perfection but the variety of possible imperfections. A perfect sphere has perfect curvature which could be distorted by being out-of-round or by a local "flat spot" or a bulge. It also has a perfectly smooth surface which could be flawed by texture or one- or two-dimensional flaws (pits and nubs or creases and folds). Consequently measurement must have a purely 3-dimensional element (volume inscribed by the center), 3-dimensional elements of local surface curvature (curvature distributions over smaller surface regions), 2-dimensional elements over the sphere (surface texture), 2-dimensional elements (maps of indented or raised curves and lines on the surface), and essentially 1-dimensional elements (aberrant points). Uncertainties are not only multi-dimensional, they are also probability distributions over the surface.

Like other complex problems, the inferences and the quantification of uncertainty from the measurement of distortion requires taking the problem apart, beginning with the most directly accessible (usually largest scale) variation, then moving to the next. An apparatus for measuring the sphere is shown in Figures 3a and b. A laser aimed at a point on the sphere is refracted onto a plate to measure the refraction angle; measurements can be replicated by shifting the plate as shown in Figure 3b. Determining the measurement points and the replication pattern is a problem in statistical experimental design.

7. Looking Forward

The mission for statistics in science agencies remains the same: to enable science. That means demonstrating to scientists and to agency administrations that science is better off *with* statistics as an integral part than without.

The first challenge to statistics as a discipline and a profession is to embed statistical thinking and formalism into BIG Science. This requires integrating stochastics into complex models and using statistical analysis to replace "brute force" computation by nonstatisticians. The second challenge is to rediscover the canonical statistical issues. These include probability sampling in the presence of nonhomogeneity, clumping, and compound processes that generate data, also experimental design for "undoable" experiments whether these must be simulated or partially designed because they are partially controllable or whether they are dynamic processes where the design can only be controlled conditional upon a sequence of results. The issue of inference and variability (uncertainty) lies at the heart of statistics. Now these issues must be addressed when both statistical and nonstatistical components are interlinked. Also both theory and methodology for these fundamental notions must be expanded to accommodate empirical data bases as the source of probabilistic representations and probability statements.

Therefore the success of statistics in the physical science agencies depends on scientifically astute statisticians and training of the next generation through vigorous

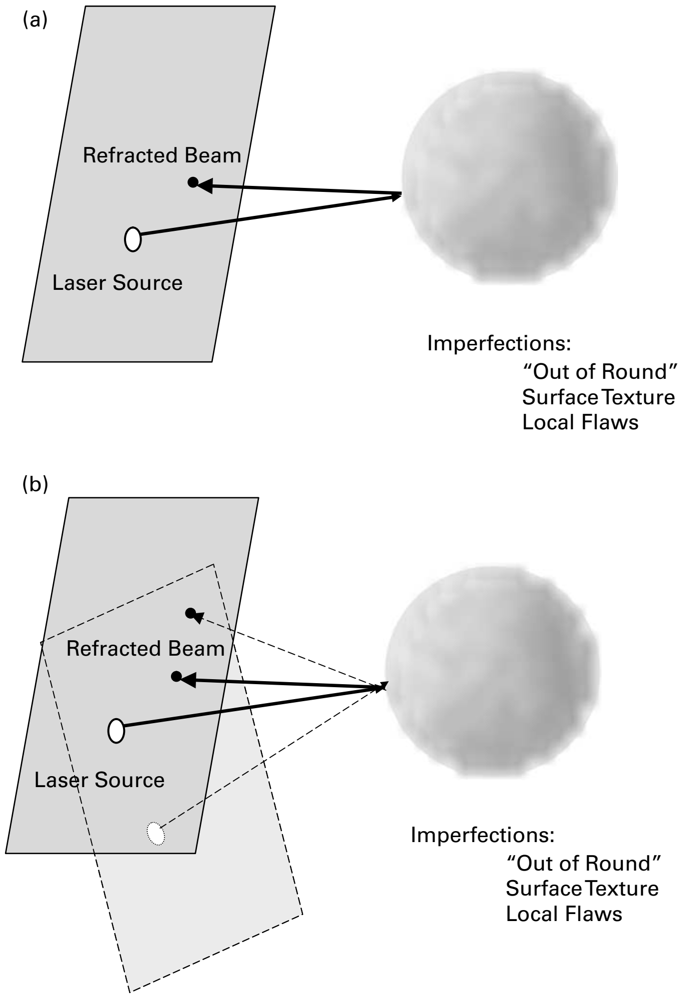

Fig. 3. Measuring Sphericity: (a) Laser Refraction Angle (b) Replication

post-doc programs, since doctoral study is too short to learn theory, research tools, implementation *and* science. The way forward also depends on enabling science in visible and tangible ways, redefining the questions that Statistics considers "answerable" and transporting methodology across applications. Respect for statistics depends on scientists and others recognizing that statisticians do offer something better for science than the quantitatively-minded scientist can achieve independently.

The comparative advantage of statistics rests in the ability to achieve greater precision with fewer errors and a deeper understanding.

8. Success in a Statistics Career

The ideal candidate to be a statistician in a science agency has a love of science, probably a particular science, and the eagerness to keep learning. Statistical specialization might be in almost any area, but a statistician also needs breadth in methodological expertise to be able to apply different statistical tools to different aspects of a scientific investigation. Methodological, modeling and computational expertise are all highly sought to meet a broad range of challenges.

The fundamentals of acquiring data, including sampling for complex, nonhomogeneous cases and experimental design for high-dimensional problems, simulations, black box prediction/modeling, provide the ability to apply the statistical principles to highly nonstandard cases and to contribute efficiency to the scientific investigation. As the science has become almost universally high-dimensional with multivariate outcome data, the applicable tools come from spatial statistics, functional data analysis and high-dimensional data analysis (data mining).

Computing skill for complex models involves numerical methods and algorithms. Computational challenges abound for massive data and for high-dimensional data with complex internal structures such as linked network-type data. Data representation is increasingly graphical and dynamic. Specialized statistical software developed to translate methodology to specific applications can provide quality control with real-time analysis and modeling, and hand "field-ready" software to scientists.

In short, career success comes from successful science: enabling science, understanding the sources of uncertainty for the inferences and expanding the range of "answerable" questions. Communication, written and verbal, is essential to articulate concepts, assumptions, results, interpretations in three different languages: in the vernacular of statistics, in scientific language and in "lay language." "Written" communication also includes data representation, especially in the form of statistical graphics.

There are two myths about careers in science agencies. The first is that success does not depend on publications; the second is that success does not involve teaching. Both are wrong. Science moves forward through refereed scientific publications, and new methodology is validated through refereed statistical publications. So, technical publications in both scientific and statistical journals are used as measures of the success of the science and of the statistical contribution. Teaching may not happen in a classroom, but as part of a multidisciplinary team. For the statistician as articulator of the principles of statistical design and uncertainty assessment, teaching takes place continually in the form of mutual education of collaborators. Indeed the statistician's success depends on it.

9. References

Clement, T.S., Hale, P.D., Williams, D.F., Wang, C.M., Dienstfrey, A., and Keenan, D.A. (2006). Calibration of Sampling Oscilloscopes with High-speed Photodiodes. IEEE Transactions on Microwave Theory and Techniques, 54, 3173–3181.

Filliben, J.J. (2005). Statistical Approaches in the NIST World Trade Center Analysis. Proceedings of the 9th International Conference on Structural Safety and Reliability, Rome, June 19–23.

Hale, P.D., Wang, C.M., Williams, D., Remley, K., and Wepman, J. (2006). Compensation of Random and Systematic Timing Errors in Sampling Oscilloscopes. IEEE Transactions on Instrumentation and Measurement, 55, 2146–2154.

Natrella, M.G. (1963). Experimental Statistics, NBS Handbook 91. National Bureau of Standards, Washington, DC. Republished (1966 & 1984). Wiley Interscience, New York. Republished (2005). Dover Publications, Mineola, NY.

NIST (2005). NIST NCSTAR 1: Federal Building and Fire Safety Investigation of the World Trade Center Disaster: Final Report of the National Construction Safety Team on the Collapses of the World Trade Center Towers. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD.

Sedransk, N. (2007). Measuring Objects and Holes. Invited Address to Louisiana Chapter of American Statistical Association Annual Meeting, November 30.

Statistical Engineering Division (2003). Statistical Engineering Division 2002 Report of Activities. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD. Also at www.itl.nist.gov/div898/pubs/ar/SED2002.pdf

Statistical Engineering Division (2004). Statistical Engineering Division 2003 Report of Activities. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD. Also at www.itl.nist.gov/div898/pubs/ar/SED2003.pdf

Statistical Engineering Division (2005). Statistical Engineering Division 2004 Report of Activities. National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD. Also at www.itl.nist.gov/div898/pubs/ar/SED2004.pdf

Wang, C., Coakley, K.J., Hale, P.D., and Clement, T.S. (2002). Uncertainty of Oscilloscope Timebase Distortion Estimate. IEEE Transactions on Instrumentation and Measurement, 51, 53–58.

Williams, D., Remley, K.A., Hale, P.D., Wang, C.M., and Clement, T.S. (2006). Sampling-Oscilloscope Measurement of a Microwave Mixer with Single-digit Phase Accuracy. IEEE Transactions on Microwave Theory and Techniques, 54, 1210–1217.

Received August 2010