# Statistical Analysis of Network Analyzer Measurements

John R. Juroshek
Electromagnetic Fields Division
NIST

C. M. Wang
Statistical Engineering Division
NIST

George P. McCabe
Department of Statistics
Purdue University

For the past five years, the Automatic RF Techniques Group (ARFTG) has conducted a measurement comparison program for vector network analyzers. Five traveling verification kits have been in circulation for measurement by participating laboratories. The accuracy of those measurements varies substantially, and classic statistical measures such as the mean and standard deviation are significantly distorted by a few participants whose measurements differ significantly from the others. This report describes some robust statistical techniques for analyzing those measurements. The techniques described are based on calculating the deviation from the median, and they are not unduly influenced by outliers or bad data. The performance of each participant is summarized by three numbers: the mean deviation, and the 10th and 90th percentile deviations.

Key Words - Vector network analyzer, scattering parameters, verification procedures, laboratory intercomparison, microwave measurements, coaxial measurements.

## Introduction

Modern vector network analyzers can make hundreds of measurements in minutes. The accuracy of these measurements is dependent not only on the accuracy of the network analyzer's electronic and microwave hardware, but also on other external factors. Vector network analyzers are typically calibrated daily, and the accuracy of their measurements after calibrations can vary substantially depending on the operator's skills, and the condition of the test ports, connectors, and cables that are used in the calibration. Assessing the accuracy of a network analyzer is a difficult, multidimensional problem.[1, 2]

For the past five years, the Automatic RF Techniques Group has conducted a measurement comparison program for vector network analyzers.[3] The purpose of this program is to help participants assess the accuracy of their measurements by comparing their measurements to those of other laboratories. Five traveling measurement kits are currently in circulation for GPC-7, Type-N, 3.5 mm, 2.92 mm, and 2.4 mm connector types. The participant measures the devices in these kits and then sends the results to the National Institute of Standards and Technology (NIST) for analysis.

NIST serves as the pilot laboratory and is responsible for maintaining the data base and for analyzing the data. The measurements are analyzed automatically, and the participant normally receives a summary report within a week or two after sending the data in for analysis.

The number of participants that have measured the kits is shown in the table below. Also shown in the table is the maximum frequency for each of the kits.

| Connector Type | Participants | Max. Frequency, GHz |
|---|---|---|
| GPC-7 | 28 | 18 GHz |
| Type-N | 38 | 18 |
| 3.5 mm | 26 | 26.5 |
| 2.92 mm | 24 | 40 |
| 2.4 mm | 13 | 50 |

Summary of ARFTG Measurement Comparison Program.

Measurements are made from 0.1 GHz to the maximum frequency. The GPC-7 and Type-N kits are measured in frequency increments of 100 MHz, while the other kits are measured in frequency increments of 200 MHz. There are 5 devices in the kits: a 20 dB attenuator, a 40 or 50 dB attenuator, an air line, a mismatch air line (Beatty standard), and an offset short. The participant measures each device three times with a disconnect and reconnect of the device between each measurement. The mean of the 3 repeats is taken to be the participant's response in this analysis.

The analysis is complicated by the fact that each participant produces a large amount of data. For example, a single participant generates 3000 complex numbers for each 2.4 mm 2-port that is measured. The uncertainties in the measurements are typically frequency sensitive and increase with increasing frequency. Analyzing the data

at a few specific frequencies is normally unsatisfactory since a participant's measurements can look good at any given frequency and bad at others nearby. Because of the large amounts of data and the quick response times, automated analysis of the data by a computer is mandatory.

Automated data analysis also introduces some unique problems. The variability of the data is substantial. Variability occurs for a number of reasons such as stress and misalignment of the connectors on the calibration devices. Also, the experience and capability of each participant is different. Generally, the measurements of most participants agree to within 1% or better, while a few participants differ by 10% or more.

The classic statistical measurements of average and standard deviation are substantially distorted by these few participants. Initially, the intent was to monitor the data manually and remove bad data from the analysis. However, that approach quickly became impractical as the number of measurements grew, and the identification of outliers became more subjective.

In this article, we present some robust analysis techniques that can be used to assess and compare the measurement capability of each participant. These techniques are not unduly influenced by outliers or bad data. The techniques are demonstrated in Section 2 by analyzing the measurement data for a GPC-7, 20 dB attenuator and a GPC-7 offset short. Section 3 summarizes the data analysis for all the connector types.

## Data Analysis for GPC-7 Devices

Two variables associated with the $S_{11}$ parameter are analyzed: the magnitude $|S_{11}|$ and the phase $ARG(S_{11})$. The magnitude and phase are analyzed separately since that is the way most participants view and use the data. Also, the error mechanisms in a vector network analyzer can be different for both magnitude and phase.

Figure 1 displays $|S_{11}|$ as a function of frequency for all participants. It shows that most of the participants are in good agreement over the frequency range. Figure 2 shows the minimum, the 25th percentile, the 50th percentile (median), the 75th percentile, and the maximum of the 28 responses at each frequency. It indicates that the middle 50% of the measurements (between the 25th and the 75th percentiles) is contained within a narrow band. Both figures show the presence of outliers. The mean and standard deviation are seriously distorted by the few outliers. In these circumstances, robust statistics such as the median and the median absolute deviation (MAD) are strongly preferred. [4, 5]

The MAD estimate of n observations $y_1, y_2, \ldots y_n$ is obtained as

$$1.4826 * med(|y_1-m|, |y_2-m|, \ldots, |y_n-m|) \qquad (1)$$

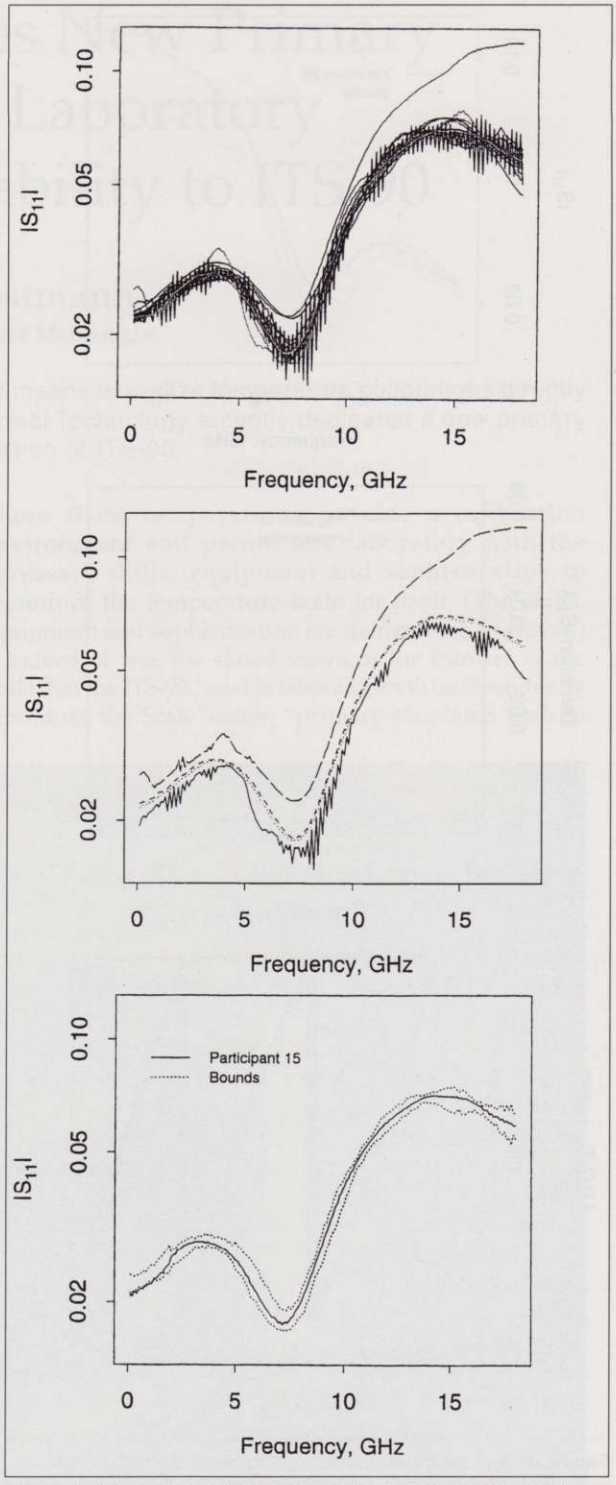


Figure 1. $|S_{11}|$ versus frequency for all participants. (top)
Figure 2. Minimum, 25th, 50th, 75th percentiles, and maximum $|S_{11}|$. (middle)
Figure 3. $|S_{11}|$ and bounds for participant 15. (bottom)

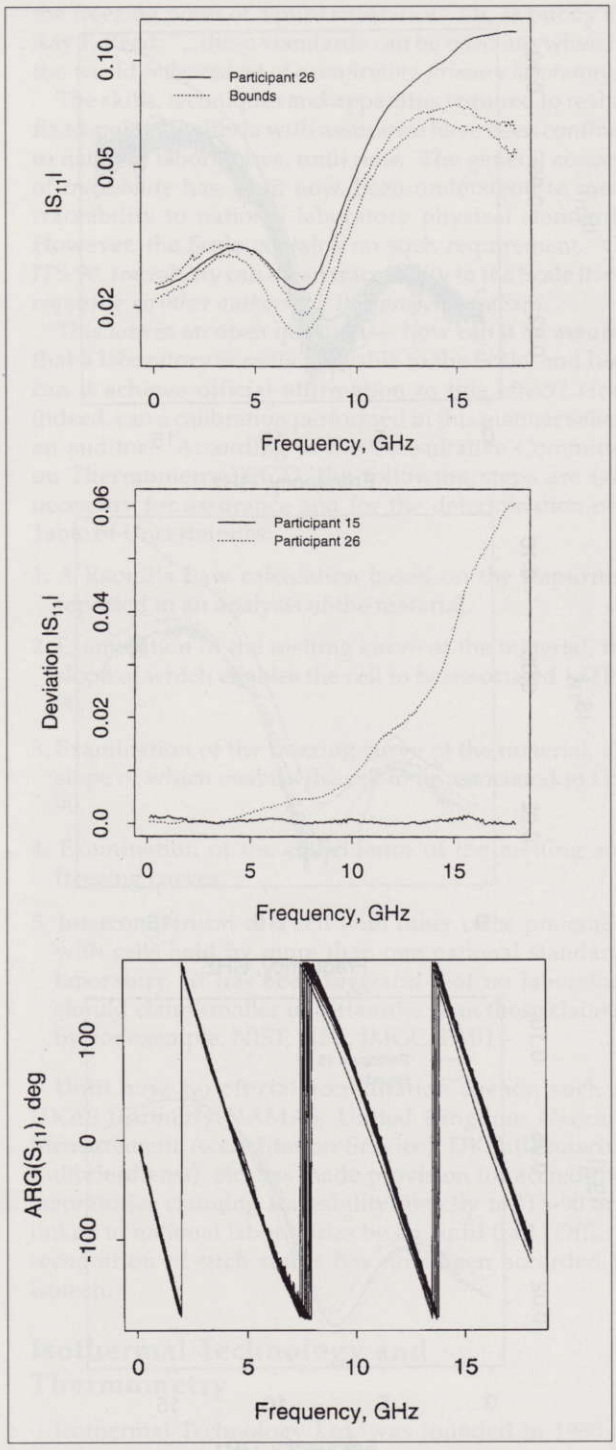


Figure 4. $|S_{11}|$ and bounds for participant 26. (top)

Figure 5. Absolute deviations from the median in $|S_{11}|$ for participant 15 and 26. (middle)

Figure 6. ARG ($S_{11}$) versus frequency for all participants. (bottom)

where the function med (.,...,.) returns the median of its arguments, and m is the median of $y_1, y_2, \ldots y_n$, that is, $m=\text{med}(y_1, y_2, \ldots, y_n)$.

The constant 1.4826 makes the MAD estimate consistent with a standard deviation when the data are normally distributed. The MAD estimate is robust against outlying observations. In fact, it will still be a good measure even when almost 50% of the observations are outliers.

The median and the MAD of the 28 $S_{11}$ magnitudes are calculated at each frequency. We use the median ±2 x MAD as the bounds for comparing the performance among participants. If the measurement capability of a participant is comparable to the others, the plot of the measurements is likely to lie within these bounds, which is the case for participant 15 as shown in Figure 3. The dotted lines are the lower and upper bounds, while the solid line is the measured magnitude for participant 15. For participant 26, shown in Figure 4, the magnitude measurements are clearly high for frequencies above 5 GHz.

The absolute deviation from the median is used to quantify the measurement differences between participants 15 and 26 as displayed in Figures 3 and 4. Let $y_{ij}$ be the measured $|S_{11}|$ for participant i at frequency j, where i=1, 2, ..., 28, j =1, 2, ..., 180, and $m_j$ is the median at frequency j. The absolute deviation from the median for participant i at frequency j is defined as

$$d_{ij} = |y_{ij} - m_j| \qquad (2)$$

Figure 5 plots $d_{ij}$ for both participants 15 and 26. The measurement differences in $|S_{11}|$ is evident in the plot.

The performance of each participant is characterized by the distribution of the absolute deviations from the median. This distribution is computed for all deviations at all frequencies. We have found that a three-number summary of the distribution is useful in describing the deviation of a participant. These numbers are the mean, and the 10th and 90th percentiles. The mean deviation is used instead of the median since a participant is normally interested in any large deviations that may exist. The 10th and 90th percentiles are useful in describing the variation and the extremes of the distribution. The three-number summaries for participant 15 is (0.00011, 0.00057, 0.00109) and (0.00029, 0.01575, 0.04983) for participant 26. The three number summary characterizes a participant's performance over the entire frequency band.

Automated analysis of phase data presents some unique problems for phase measurements near ±180°. For example, the mean of the two phase measurements +179.5°and −179.5° is zero, even though both measurements are close to ± 180°. Computing the mean and standard deviation of phase data with conventional statistical methods can produce misleading results since

the measurements are circular at ± 180°. The following techniques avoid these problems.

Again, the mean of the 3 repeats at each frequency is used as the participants response in the analysis. The mean for the n angular data is obtained as θ1, θ 2, ..., θ n is obtained as: [6]

$$\tan^{-1}\left(\sum_{i=1}^{n}\sin\theta_i \Big/ \sum_{i=1}^{n}\cos\theta_i\right) \qquad (3)$$

Figure 6 is a linear plot of the mean phase as a function of frequency for all participants. Using a linear plot to display phase data is not adequate for comparing phase among participants, particularly in the neighborhood of ± 180°. A polar plot can also be used, but it also is not a good tool for comparing the phase, since it is difficult to tell the magnitude of the difference from those plots.

The difference between 2 angles $\theta_1$ and $\theta_2$ (in degrees) can be described as: [6]

$$d(\theta_1, \theta_2) = 180 - |180 - |\theta_1 - \theta_2|| \qquad (4)$$

This difference is always between 0° and 180°, and is no longer circular. With this technique it is easier to work with the "differenced" phase data. The phase response for each participant can be compared to either the mean or the median of all participants. To be consistent with the preceding analysis, we use the median. At each frequency, we calculate the median of the phase for the 28 participants. The median for the n angular $\theta_1$, $\theta_2$, ..., $\theta_n$ is obtained using

$$\tan^{-1}\left(med(\sin\theta_1, \ldots, \sin\theta_n) \,/\, med(\cos\theta_1, \ldots, \cos\theta_n)\right) \qquad (5)$$

The median is then subtracted from the phase angle data for each participant using the circular difference defined above. With this technique it is possible to plot and compare the resulting phase (deviation from the median) among participants. Figure 7 displays (solid line) the phase deviation for participant 15. It shows that the deviations are close to 0°, indicating a good agreement with the median, while for participant 8, shown in Figure 8, the deviations are as large as 10° at some frequencies. The dotted line in both plots is the upper bound (2 x mad) for the phase deviation based on all 28 participants. "Consistent" participants should have their phase deviations below this upper bound. The bound also shows that the variation in phase measurement among participants is largest near 8 GHz. The reason for this is that the device has $|S_{11}|$ near 0 at this frequency. The uncertainty in phase measurements on a vector network analyzer increases with decreasing $|S_{11}|$. Again, we use the average, the 10th percentile, and the 90th percentile of the phase deviation to characterize the participant's capability in measuring the phase.

The analysis described here is based on calculating the

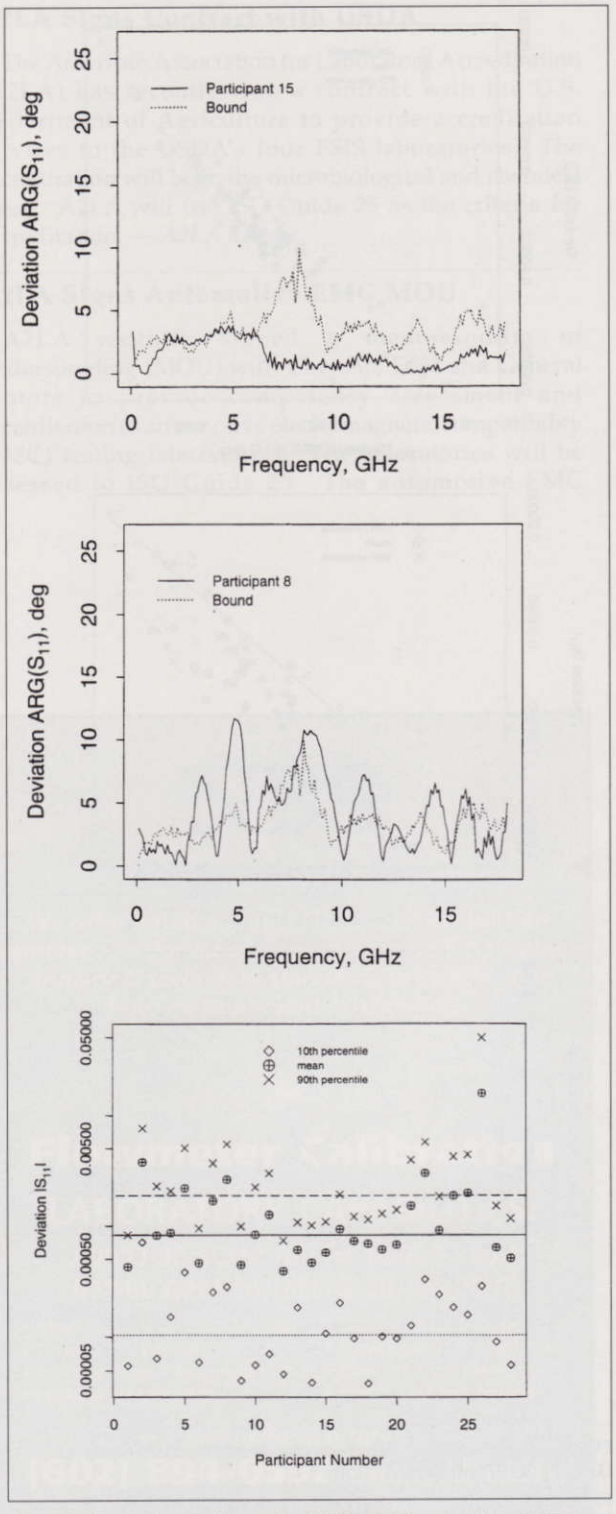


Figure 7. Deviation and bound in ARG ($S_{11}$) for participant 15. (top)
Figure 8. Deviation and bound in ARG ($S_{11}$) for participant 8. (middle)
Figure 9. Deviation in $|S_{11}|$ for a GPC-7, 20 dB attenuator. (bottom)

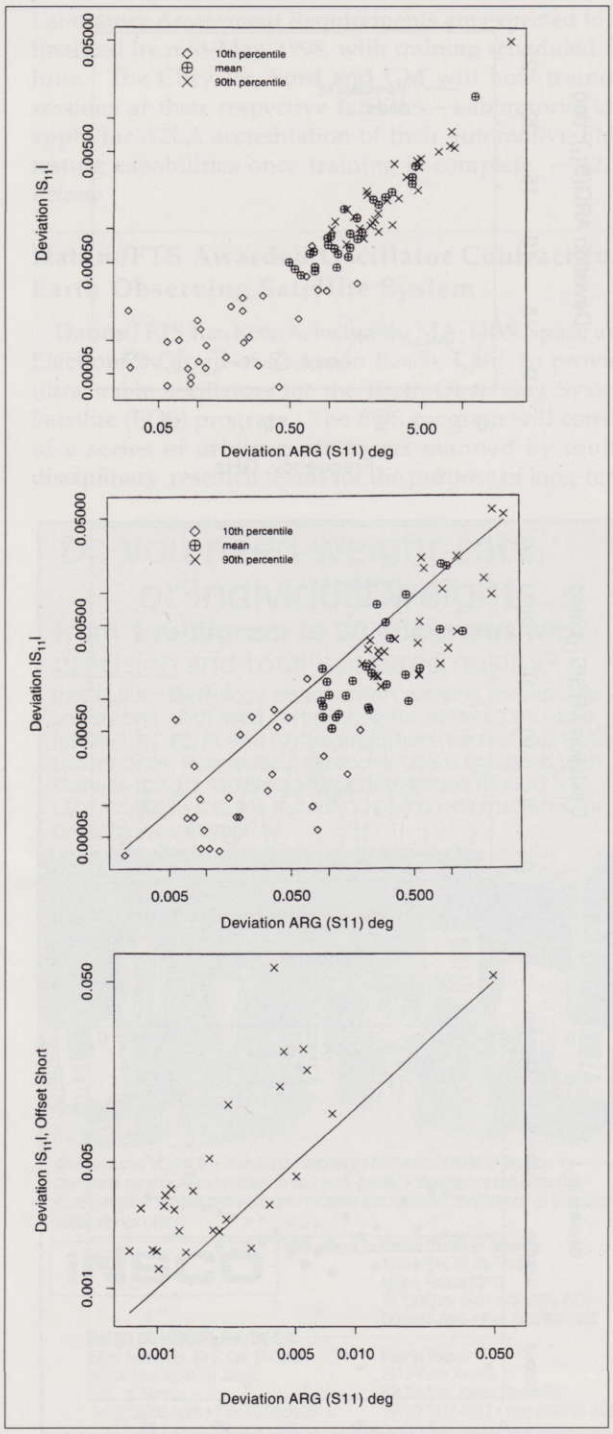


Figure 10. Deviation in $|S_{11}|$ versus deviation in ARG ($S_{11}$) for a GPC-7, 20 dB attenuator. (top)

Figure 11. Deviation in $|S_{11}|$ versus deviation in ARG ($S_{11}$) for a GPC-7, offset short. (middle)

Figure 12. Deviation in $|S_{11}|$ for a GPC-7 offset short versus deviation in $|S_{11}|$ for a GPC-7, 20 DB attenuator. (bottom)

deviation from the median, with the assumption that the median reflects a more accurate measurement than that of any individual participant. However, that is only an assumption, and the median could be biased with systematic errors that are common to all of the measurements.

In Figure 9, the 10th percentile, average, and 90th percentile deviation in $|S_{11}|$ for the GPC-7, 20 dB attenuator are displayed for each of the 28 participants. The "typical" deviation which is the median value of the deviation for all of the participants is also displayed. The typical values of the 10th percentile, average, and 90th percentile deviations are indicated by the dotted, solid, and dashed lines, respectively. This plot shows the variability among the participants. The 10th percentile deviation has less discriminating power, and its primary use is to show the best that can be expected. In some cases the 10% deviation is nearly equal to the number of significant figures in the reported data. The spread among the participants is typically more than one order of magnitude.

Figure 10 shows the deviation in $|S_{11}|$ for the GPC-7, 20dB attenuator, plotted as a function of the deviation in ARG($S_{11}$). This plot is useful for characterizing the participant's deviation in both magnitude and phase. Figure 10 indicates that participants with large magnitude deviation usually also have large phase deviation. The correlation coefficient for the mean magnitude and phase deviations is 0.973.

Figure 11 shows a similar plot for the GPC-7 offset short. The deviation in $|S_{11}|$ is plotted versus the deviation in ARG($S_{11}$). If the measurement error vector has a circularly symmetrical probability distribution, then the points should fall along a line described by

$$D_M = \pi D_{ARG} / 180 \qquad (6)$$

where $D_M$ is the deviation in $|S_{11}|$ and $D_{ARG}$ is the deviation in ARG($S_{11}$) in degrees. Equation (6) is simply the small angle approximation formula that relates the phase angle, in degrees, to the magnitude of the arc. This equation is shown as the solid line in Figure 11. As can be seen, the deviations in phase are greater than expected for a circularly symmetrical probability distribution. Theoretical studies of the errors in RF connectors have shown that the deviation in phase (in radians) for an offset short can be 4 times greater than the deviation in magnitude [7]. The results shown in Figure 11 support this theory.

Figure 12 plots the 90th percentile deviation in $|S_{11}|$ for an offset short versus the 90th percentile deviation in $|S_{11}|$ for a GPC-7, 20 dB attenuator. If the deviations were equal for both devices, the data would fall along the solid line shown in the figure. The deviation in $|S_{11}|$ for the offset short is generally larger than for the 20 dB

attenuator. That result is expected since the network analyzer's residual source match and the residual reflection tracking are significant error sources for measurements of high reflection devices, but not for low reflection devices.

Figure 13 plots the deviation in $|S_{12}|$ in decibels, for the GPC-7, 20 dB attenuator, versus the deviation in $ARG(S_{12})$. If the error vector has a circularly symmetrical probability distribution, the deviations should fall along the line

$$D_{M,dB} = 8.69\,\pi\,D_{ARG} / 180 \qquad (7)$$

where $D_{M,dB}$ is the deviation in $|S_{12}|$ in dB. Equation (7) is obtained by converting $D_M$ in (6) to decibels. This equation is plotted as a solid line in Figure 13. Again, the deviation in phase is larger than expected for a circularly symmetric probability distribution of the error vector.

## Summary of Measurements for Other Connectors

Figures 14 through 18 summarize the results for all of the connector types. In each of these plots, the 10th percentile, average, and 90th percentile deviations are displayed for all participants and all five connector types. The typical values of the 10th percentile, average, and 90th percentile deviations are also shown by dotted, solid, and dashed lines respectively.

Figure 14 shows the deviations in $|S_{11}|$ for the 20 dB attenuator in each of the 5 kits. Surprisingly, the typical deviation is nearly equal for all 5 connector types, even though the smaller connector sizes are measured at higher frequencies. Initially, the expectation was that the performance of the smaller size connectors would degrade because of their small size, and higher operating frequency. However, there is not a strong indication of that trend in the present data.

Figure 15 shows the deviations in $|S_{11}|$ for the offset shorts, while Figure 16 shows the deviations in $ARG(S_{11})$ for the offset shorts. In Type-N, the phase deviation for some of the participants is particularly large due to a confusion in identifying the identification of sex of the test ports and the device under test.

Figure 17 shows the deviations in $|S_{12}|$ for a 20dB attenuator, while Figure 18 shows the deviations in $ARG(S_{12})$ for that device.

## Conclusions

The analysis of network analyzer data is complicated by the large amounts of data that these instruments produce. For example, a single participant generates 3000 complex numbers for each 2-port device that is measured

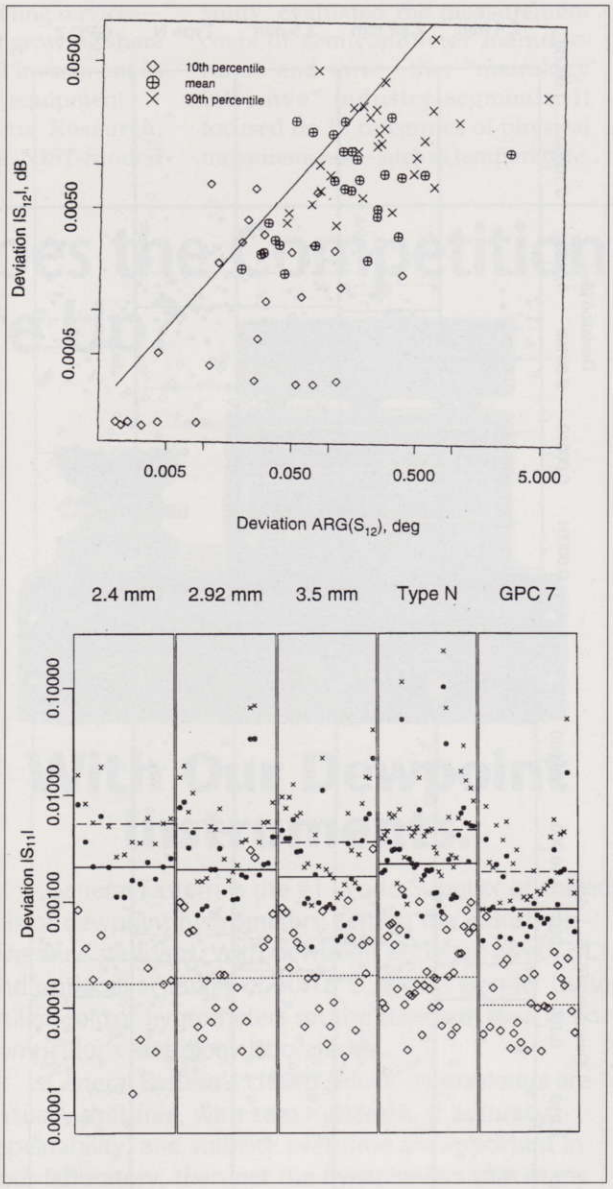


Figure 13. Deviation in $|S_{12}|$ versus deviation in ARG $(S_{12})$ for a GPC-7, 20 dB attenuator. (top)

Figure 14. Deviation in $|S_{11}|$ versus connector type for all 20 dB attenuators. (bottom)

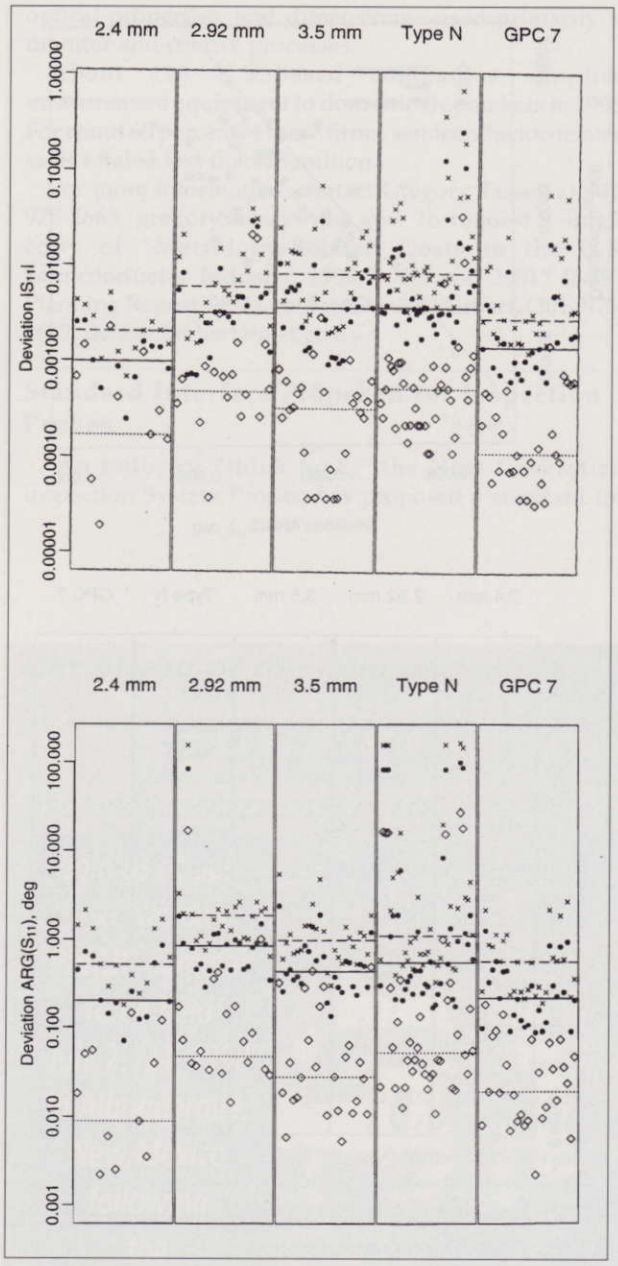


Figure 15. Deviation in $|S_{11}|$ versus connector type for all offset shorts. (top)

Figure 16. Deviation in ARG ($S_{11}$) versus connector type for all offset shorts. (bottom)

in the 2.4 mm traveling kit. The accuracy of the measurements varies substantially with frequency. The experience to date indicates that a number of participants are making reflection measurements to within 1 % of the median or better. However, there are also a number of participants whose deviation from the median is 10% or greater.

Participants with large deviations substantially bias the average. Eliminating outliers or bad data from the data base became impractical as the number of measurements grew, and the identification of bad data became more subjective. This article describes some robust techniques for analyzing network analyzer measurements. The analyses are based on calculating the deviation of each participant's measurements from the median. The techniques described here are reasonably insensitive to outliers and bad data. Three measures of deviation have been found to be particularly useful in summarizing the performance of a participant. They are the 10th percentile, average, and 90th percentile deviations from the median.

The "typical" 10th percentile, average, and 90th percentile deviations in the measurements are surprisingly similar for all of the connector types. Initially, the expectation was that the performance of the smaller connector sizes would be degraded due to their smaller size, and higher operating frequency. However, that trend was not observed in the current data.

---

References

1. Ide, J.P., K. Hilty and J.P.M. de Vreede, "Trilateral International Type-N ANA Comparison Exercise," Metrologia, 32,1995, pp 3541.

2. Ridler, N.M., and J.C. Medley, "A Comparison of Complex Scattering Coefficient Measurements in 50 Ohm Coaxial Line to 26.5 GHz," NPL Report DES 138, National Physical Laboratory, Teddington Middlesex, UK, June 1995, pp 140.

3. Judish, R.M., and J.G. Burns, "Measurement Program Compares Automatic Vector Network Analyzers," Microwaves & RF, May 1991, pp 203-206.

4. Hampel, F.R., Roncheti, E.M., Rousseeuw, P.J., and Stahel WA., *Robust Statistics: The Approach Based on Influence Functions*, Wiley: New York, 1986.

5. Velleman, P.F., and Hoaglin, D.C., *Applications. Basics and Computing of Exploratory Data Analysis*, Duxbury: Boston, 1981.

6. Mardia, K.V., *Statistics of Directional Data*, Academic Press: New York, 1972.

7. Juroshek, J.R.,"A Study of Measurements of Connector

Repeatability Using Highly Reflecting Loads", IEEE Trans. MTT, Vol. MTT—35, No.4, April 1987, pp 457-460.

---

John R. Juroshek and C.M. Wang are with the Electromagnetic Fields Division and the Statistical Engineering Division of the National Institute of Standards and Technology, Boulder, CO 80303. Tel 303-497-5362, juroshek@boulder.nist.gov.

George. P. McCabe is with the Department of Statistics, Purdue University, West Lafayette, IN, USA.



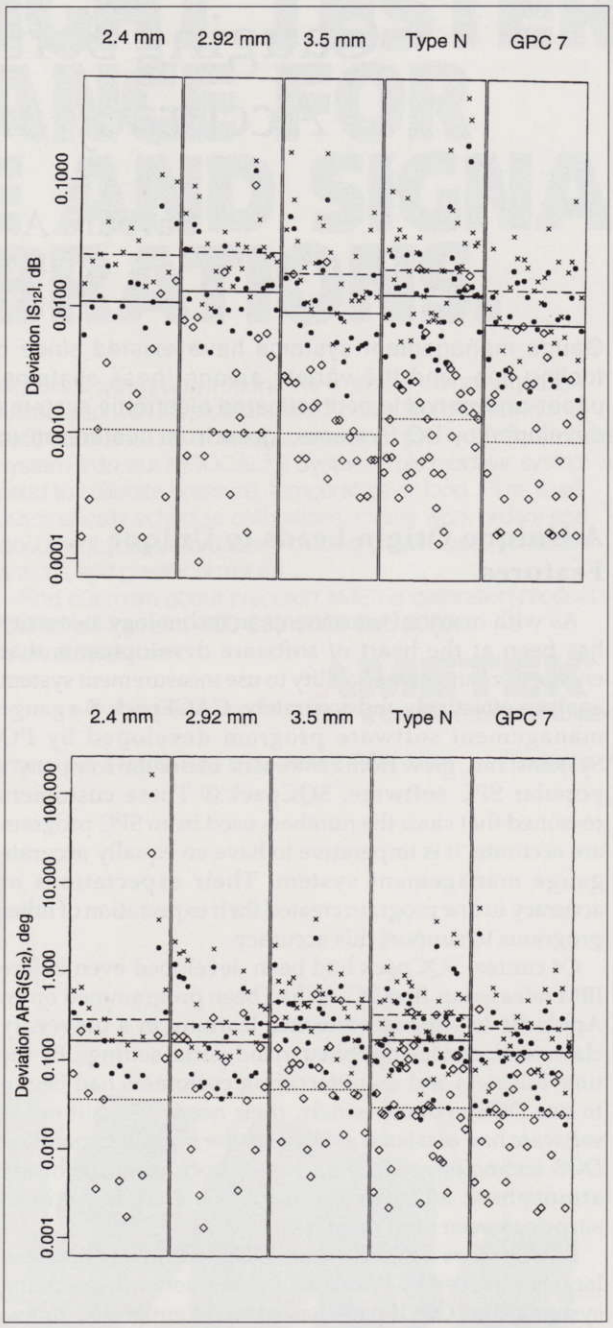


Figure 17. Deviation in $|S_{12}|$ versus connector type for all 20 dB attenuators. (top)

Figure 18. Deviation in ARG ($S_{12}$) versus connector type for all 20 dB attenuators. (bottom)