A Statistical Approach to Measurement Uncertainty

MICHAEL D. FOEGELLE
EMC Test Systems, Austin, TX*

Despite the fact that most authors have skirted the subject of Type A analyses, the method is not so difficult that it should be avoided altogether.

Introduction

Over the past couple of years, the topic of measurement uncertainty has come to the forefront in the international EMC community. In brief, the intention of measurement uncertainty is to take the more traditional terms of precision, accuracy, random error, and systematic error used in scientific circles and replace them with a single term. This term represents the total contribution to the expected deviation of a measurement from the actual value being measured.[1-2]

In general, precision is a measure of random error, or how closely repeated attempts hit the same point on a target, while accuracy is a measure of systematic error, or how close those attempts are to the center of the target. It is obvious that both contributions must be accounted for in order to determine the quality of a measurement, although the combination of the two can sometimes be more subtle than might be expected. There has been some discussion over whether the term reproducibility is also replaced by uncertainty since the concept of reproducibility must contain variations in the equipment under test (EUT) and therefore does not represent the same quantity as measurement uncertainty.[3]

The methods for determining a measurement uncertainty have been divided into two generic classes:

- Type A represents a statistical uncertainty based on a normal distribution.
- Type B represents uncertainties determined by any other means.

In last year's ITEM, Manfred Stecher wrote an article describing the introduction of uncertainty evaluations into various EMC standards and explained the technique typically used to determine measurement uncertainties for EMC measurements.[4] (A similar paper was also presented at the 1996 IEEE International EMC Symposium in Santa Clara, CA.[5]) The article gives an adequate introduction to the Type B evaluation method, which uses individual measurements, manufacturers' specifications, and even educated guesses to determine a combined uncertainty. However, the author is a little too quick to discard the statistical Type A uncertainty measurement as impractical. To be sure, the Type A analysis does suffer from the very pitfalls which Mr. Stecher points out. However, with a bit of care it is possible to obtain a significant amount of useful information from the technique.

The advantage of a Type A uncertainty measurement is that when done correctly, the resulting value is irrefutable since it has been determined from real world measurements. The biggest complaint I hear from engineers being exposed to the Type B uncertainty budget method for the first time is the fact that too many of the terms are either poorly defined by equipment manufacturers or must simply be estimated. In many cases, the chosen values may be too stringent in order to provide a safety margin. On the other hand, the desire for smaller total uncertainties can lead to using smaller estimations than is realistic for some terms that are hard to determine.

Antenna manufacturers have easy access to a vast database of antenna calibrations with which to determine statistical trends. However, as the data shown here will demonstrate, it is not necessary to have an extremely large sample to get acceptable results. The real issue in using a statistical approach is in determining where it fails and using a Type B analysis to fill in the gaps. The document NIS 81, "The Treatment of Uncertainty in EMC Measurements," released by NAMAS,[1] recommends this exact approach.

Certainly a Type A analysis of a set of measurements can be expected to include all random errors of the entire measurand and none of its systematic errors. But does that mean that the Type A uncertainty contains only the random portion of the individual terms which might be used in a Type B analysis? Certainly not. It is not possible to completely separate ran-

 *See advertisement on page 1

dom and systematic effects into individual categories.[2] Random effects such as positioning error[6] or cable length and frequency dependence of standing waves can serve to randomize such systematic errors as site imperfections or mismatch errors. Intentionally varying the setup between repeated measurements can do the same.

The discussion presented here will focus primarily on antenna calibration measurements, which have many similarities to radiated emission measurements. However, many of the techniques demonstrated will be applicable to radiated susceptibility and conducted tests. Despite the fact that most authors have skirted the subject of Type A analyses, the method is not so difficult that it should be avoided altogether. In fact, many of the more difficult terms to determine for a typical Type B analysis are already included in the random error of the total measurement, thus reducing the overall task.

Type A Evaluation of Uncertainty

Random effects cause repeated measurements to vary in an unpredictable manner. The associated uncertainty can be calculated by applying statistical techniques to the repeated measurements. An estimate of the standard deviation, $s(q_k)$, of a series of n readings, q_k, is obtained from

$$s(q_k) = \sqrt{\frac{1}{(n-1)}\sum_{k=1}^{n}(q_k - \overline{q})^2}$$

where $\overline{q}$ is the mean value of n measurements. The random component of the uncertainty may be reduced through repeated measurements of the EUT. In this case, the standard deviation of the mean, $s(\overline{q})$, given by

$$s(\overline{q}) = \frac{s(q)_k}{\sqrt{n}}$$

represents the uncertainty of the resulting mean. This last point has been misinterpreted by some due to a confusing statement in section 3.2.6 of NIS 81, "The standard uncertainty, $u(x_i)$, of an estimate x_i of an input quantity q is therefore $u(x_i) = s(\overline{q})$." This statement is certainly true, but some readers have construed it to mean that the standard uncertainty of a single measurement, q_k, is given by $s(\overline{q})$.[7-8] This is true if n = 1 so that $s(\overline{q}) = s(q_k)$. This is explained more clearly in section 3.2.3 of NIS 81 where the concept of predetermination is discussed.

Time constraints and other practical considerations will often make it unfeasible to perform more than a single measurement on an EUT. However, repeat measurements can be performed on a similar EUT to predetermine $s(q_k)$ as the expected standard uncertainty of an individual measurement. If a smaller uncertainty due to random errors is desired, multiple measurements can be made and the value of $s(\overline{q})$ reduced accordingly. The value of n used in this case is the number of measurements made on the EUT, not the number of measurements used in the predetermination.

It should be noted that variations in the EUT as a function of time, as well as that of multiple like EUTs as in the method demonstrated here, will be included in the uncertainty determined through this method. For this reason, and to provide a means to determine some of the systematic contributions to the uncertainty, it is recommended that a stable reference radiating source such as a comb generator or amplified noise source be used as the EUT for uncertainty determinations. This provides a repeatable EUT which, when used in conjunction with "round robin" type testing, can allow determination of even the systematic elements of your measurement uncertainty to within the uncertainty of the round robin test. This type of EUT also provides a broad continuous frequency range for uncertainty determination as opposed to the random spectrum points of a typical EUT. If a reference source is not available, a number of the same benefits may be obtained using a stable signal generator and appropriate radiating antenna. However, transmit cable placement and other effects will add some additional random error and this method does not lend itself to inter-site comparisons.

Uncertainty Example: Antenna Calibration

Antenna calibrations have uncertainty aspects similar to that of both radiated emissions and susceptibility tests. However, the one systematic element missing from the test is the absolute value of the fields (or signal levels) involved. Thus, the test only depends on the linearity of the instrumentation and not its absolute calibration. However, this does not affect the validity of this method for determining Type A uncertainties for EMC tests since the systematic error in field level cannot be determined by this method without inter-site comparisons or other tests using multiple methods to determine the absolute field level.

To obtain the data shown here, a sample of 26 different antennas calibrated over a three-and-a-half month period was used. The antennas were identical log-periodic antennas with a frequency span of 200 MHz to 1 GHz. The time period spanned from mid-August through the end of October, providing a significant variation in temperature and weather conditions. The data were measured at 1601 frequency points using a vector network analyzer, low loss cables, and a positioning tower with 0.1 cm positioning resolution.

The test was performed at a 10-meter separation, 2-meter transmit height, and 1- to 4-meter scan height per ANSI C63.5 on an open area test site (OATS). The vector network analyzer used has over 100 dB of available dynamic range and superior external noise rejection, making it ideal for use on an OATS where ambients would be a problem for traditional spectrum analyzer/tracking generator combinations. Since the network analyzer does not

Continued on page 119

offer a max-hold or other such functionality traditionally seen on a spectrum analyzer, it is necessary to perform the max-hold by transferring individual traces to a controlling PC and allowing the test software to perform the max-hold.

To facilitate this method, it is also necessary to step the tower and take frequency sweeps at discrete heights since the network analyzer cannot sweep the entire frequency band fast enough to get acceptable height resolution when the tower moves continuously. Tests with tuned dipoles have shown that the variation in the antenna factor at 1 GHz for a 1-cm step, a 5-cm step, and a single frequency point continuous motion max-hold is less than ±0.02 dB. If the tower step size was too large, or the sweep time too slow in the case of a continuous motion, the measurement could be expected to introduce a systematic error since missing a peak signal would always result in a measurement lower than the real value. This error exists for scanned height radiated emissions tests as well as antenna calibrations. For an antenna calibration, this error would result in a larger antenna factor than is actually the case.

The standard deviation of the 26 antenna factors and their maximum deviations from the average are shown as a function of frequency in Figure 1. The negative of the standard deviation is also shown for comparison to the negative deviation. The symmetry of the positive and negative deviations is a first indication of how closely the sample approximates a normal distribution. An asymmetrical envelope would be a cause for concern and indicate the need for separate positive and negative uncertainty values at the minimum.

Neglecting, for the moment, any additional contributions to the uncertainty due to systematic errors, these data indicate that the expanded uncertainty (k = 2) of an individual calibration is less than ±0.5 dB at all frequencies. That means that an uncertainty of ±0.5 dB with greater than 95% confidence can be claimed for this calibration. A second verification of this claim is the fact that, with the exception of the negative deviation above 950 MHz, none of the antennas in the sample had antenna factors which deviated from the average more than ±0.5 dB. This provides an added level of confidence in the quality of this uncertainty value.

It should be noted that this data represents calibrations of 26 *different* antennas performed at different times. About fifty percent of the antennas were directly off the production line, but the other half were re-calibrations of antennas as many as ten years old. Although one might expect a batch of antennas to be as close to identical as possible, this sample surely must contain some amount of manufacturing uncertainty. If this manufacturing uncertainty is non-zero, then it must also be contained within the above uncertainty. Since the "perfect" antenna, in terms of manufacturing quality, is defined by the average of all antennas, this uncertainty must be totally random.

UNCERTAINTY TERMS FROM NIS-81

- Estimated standard deviation from a sample of n readings:

$$s(q_k) = \sqrt{\frac{1}{(n-1)}\sum_{k=1}^{n}(q_k - \overline{q})^2}$$

- Standard deviation of the mean of n readings:

$$s(\overline{q}) = \frac{s(q)_k}{\sqrt{n}}$$

- Standard uncertainty (of the mean of n readings) resulting from a Type A evaluation:

$$u(x_i) = s(\overline{q})$$

- Standard uncertainty for contributions with a normal probability distribution:

$$u(x_i) = \frac{U}{k},$$

where U represents the expanded uncertainty of the normal distribution (last item below).

- Standard uncertainty for contributions with rectangular probability distribution:

$$u(x_i) = \frac{a_{i+} - a_{i-}}{2\sqrt{3}}$$

for an asymmetrical distribution, where a_{i+} and a_{i-} are the bounds of the rectangular region, or

$$u(x_i) = \frac{a_i}{\sqrt{3}}$$

for a symmetrical region with bounds $\pm a_i$.

- Standard uncertainty for contributions with U-shaped probability distribution:

$$u(x_i) = \frac{a_{i+} - a_{i-}}{2\sqrt{2}}$$

for an asymmetrical distribution, where a_{i+} and a_{i-} are the bounds of the U-shaped region, or

$$u(x_i) = \frac{a_i}{\sqrt{2}}$$

for a symmetrical region with bounds $\pm a_i$ or where a_i is the larger of a_{i+} or a_{i-}.

- Standard uncertainty in terms of the measured quantity:

$$u_i(y) = c_i \bullet u(x_i).$$

- Combined standard uncertainty:

$$u_c(y) = \sqrt{\sum_{i=1}^{N} u_i(y)}$$

- Expanded uncertainty:

$$U = k \bullet u_c(y) \text{ or } U = k_p \bullet u_c(y)$$

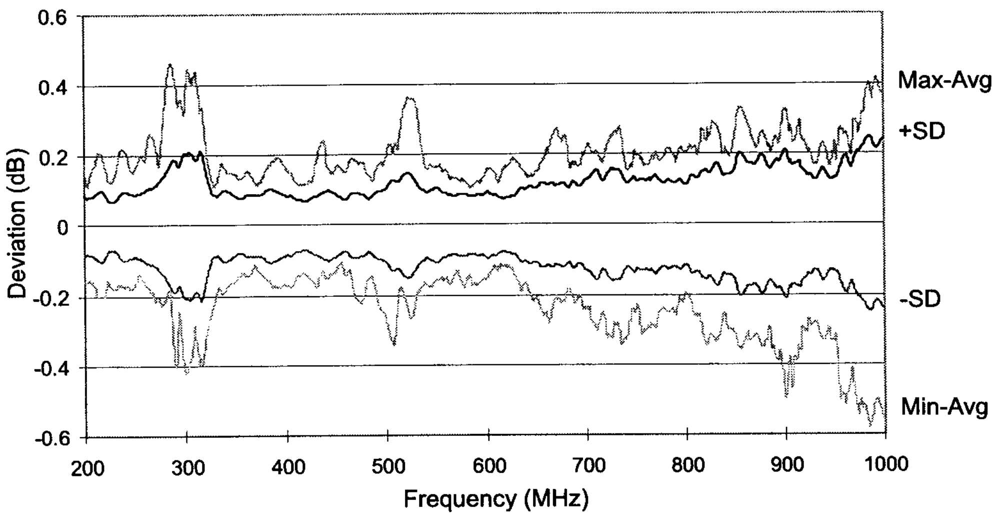

Figure 1. *The Standard Deviation, Maximum Minus Average, and Minimum Minus Average of the Antenna Factors from a Set of Twenty-six Identical Antennas. Neglecting any additional systematic errors, the expanded uncertainty (k = 2) for an individual antenna factor would then be twice the standard deviation for a value of approximately ±0.5 dB with greater than 95% confidence at all frequencies.*

Although this manufacturing may add to the total calibration uncertainty measured by this technique, as long as the resulting uncertainty is within an acceptable range, it does not matter whether or not the contribution is large or small. In this case, there is also the added benefit that the intentional introduction of totally random effects due to the antennas may help to randomize the systematic contributions to the total measurement uncertainty. For emissions measurements, this is equivalent to using multiple identical EUTs to determine the Type A uncertainty. As long as variations in the EUTs do not add excessive contributions to the random error, it is possible to obtain a suitable uncertainty value from tests of multiple EUTs.

This is not to say that multiple EUTs always provide an acceptable method for performing this type of uncertainty calculation. One or two EUTs with significant deviations from the norm can introduce a significant change to the resulting uncertainty, as shown in Figure 2. One additional antenna, a different model which varies significantly from the average at different frequencies, was introduced to the sample to demonstrate the possible effects.

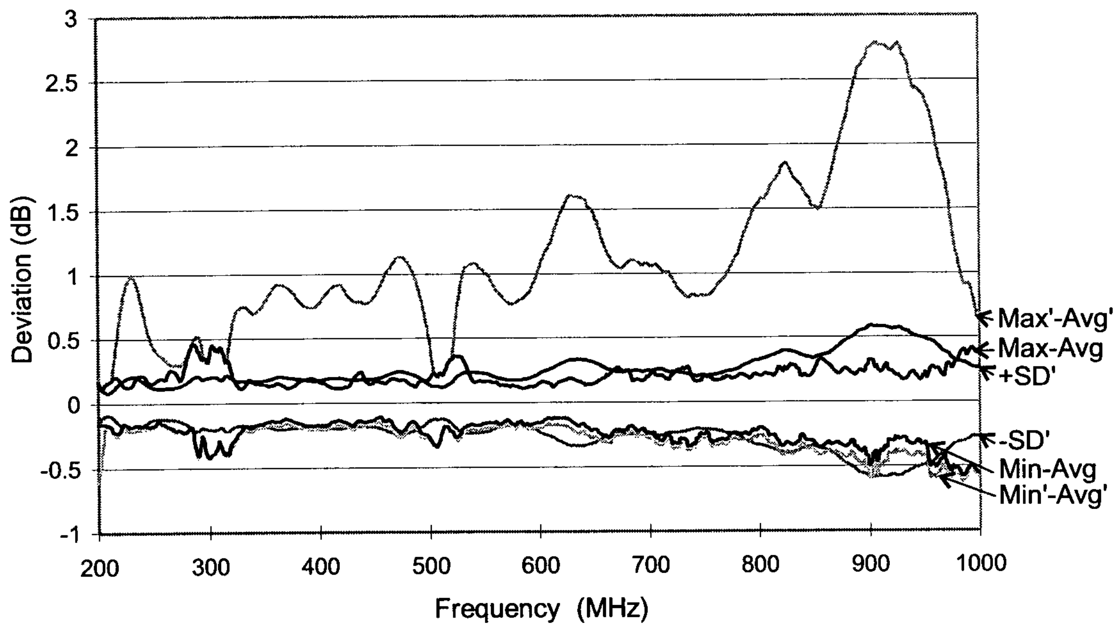

Figure 2. *The Effect of EUT Variation on the Standard Deviation. Note the Difference in the Maximum Deviation for the Primed Sample Vs. Un-primed. The effect on the standard deviation is sufficient to equal or surpass the maximum deviation in the unprimed sample.*

Note that for most of the frequency range, the deviation is below 1 dB, yet the contribution of one additional antenna is sufficient to change the standard deviation such that it is larger than the original maximum deviations at most frequencies. As this example shows, when working with relatively small samples, it is important not to introduce factors which may exaggerate the error in the measurement. Likewise, it is important to avoid arbitrarily discarding data because it makes the result look bad!

Figure 2 also demonstrates the difference between systematic errors and random errors. One sample with a large systematic error (not present in the other samples) can have a significant effect on the uncertainty, since it changes not only the standard deviation, but also the mean of the samples. This causes the negative maximum deviation (minimum - average) to be larger than before, even though the absolute value of the minimum curve is the same.

Note that the standard deviation shown in Figure 2 would still allow the claim of ±0.5 dB for the expanded uncertainty from 200 to 600 MHz. It is perfectly acceptable (and often necessary due to band breaks in equipment) to generate frequency dependent uncertainties. The expanded uncertainty from 600 to 800 MHz could then be set at ±0.8 dB and from 800 MHz to 1 GHz at ±1.2 dB.

Cleanup: Type B Evaluation

In order to develop a final value for the expanded uncertainty, it is necessary to make an evaluation of all of the typical contributions to uncertainty to determine which terms are included in the Type A result. In this case, nearly every effect exerts a random contribution due to the considerable variation in test setup and conditions over the allotted time period. Factors such as temperature and humidity effects, ground plane quality due to ground moisture, ground plane warping with temperature, test cable lengths, cable connection quality, cable calibration, etc. all varied significantly over the sample.

Table 1 lists some of the typical entries in a Type B uncertainty budget and suggests which ones are likely to be totally random, systematic, or a combination of the two. The listed distribution type represents the typical accepted distribution type for that entry. Note that most calculated values are assumed to be rectangular, which may often be a more stringent criteria than is necessary.

For the antenna calibration example given, only a few possible systematic errors may need to be accounted for. In general, a purely systematic error is likely to have a U-shaped distribution since it always represents a deviation to one side of the real value. A common example of this type of contribution is that of cable mismatch. Since a perfect match is represented by zero reflection and a mismatch in any direction results in non-zero reflection, the statistical probability that there is some mismatch causes the peaks of the distribution to occur away from the zero (matched) value. The effect of mismatch is to change the detected level by reflecting the power back towards the source. It can also generate standing waves in cables which then serve to increase or decrease the detected signal based on the frequency and cable length. A standing wave is a systematic error in a fixed system for a given frequency. However, over frequency, the effect is either constructive or destructive, giving a net random contri-

CONTRIBUTION	DISTRIBUTION TYPE	RANDOM	SYSTEMATIC	AF CAL	RE	RI
Antenna factor	Normal		X		X	
Cable calibration	Normal	X		X	X	
Coupler calibration	Normal		X			X
Receiver/probe linearity	Rectangular	X	x	X	X	X
Receiver level detection	Rectangular	x	X		X	X
Antenna directivity	Rectangular	X			X	
AF variation with height	Rectangular		X		X	
Antenna phase center	Rectangular		X		X	
Field uniformity	Rectangular	X				X
Frequency interpolation	Rectangular	X	x		X	X
Distance measurement	Rectangular	X	x	X	X	X
Height measurement	Rectangular	X	x	X	X	
Site imperfections	Rectangular	X	x	X	X	
Mismatch	U-shaped	X	x	X	X	X
Temperature effects	Rectangular	X		X	X	X
Setup repeatability	Type A	X		X	X	X
Ambient signals	Rectangular	X	x	X	X	
EUT repeatability	Type A	X			X	X

Table 1. *Various Possible Contributions to Measurement Uncertainty, along with Typical Accepted Distribution Types and Possible Contributions to Both Random and Systematic Errors. (For items which might have both random and systematic contributions, a lower-case X represents the typically smaller or less likely contribution.)*

bution. In the case of antenna calibrations, the mismatch at the antenna is a part of the calibration, so standing waves are the only contribution of concern. All other cable contributions are included in the cable calibration, which is part of the random error contribution.

Another systematic error contribution which generates a U-shaped distribution is the max-hold height step. For continuous height scanning, this is a function of sweep time versus tower speed. Since the "correct" value is a maximum, any deviation from the correct value can only be less than the correct value. As mentioned previously, this error was verified to be less than ±0.02 dB. This error is a good example of the difference between performing a Type B analysis of an entire test setup or using a mixed Type A and Type B analysis. The height step has both random and systematic components, but the random components would be measured in a Type A analysis. Thus the contribution to the mixed Type B analysis is not the same as that for the Type B-only analysis.

The final and most troubling contribution is that of site imperfections. Fortunately, several factors mitigate this contribution somewhat. Temperature changes over the test period caused dimensional changes in the surface of the metal ground plane which would have randomized the effects somewhat. Also, similar to the effects of standing waves, the imperfections in the ground plane will have different effects at different frequencies. Variations in antenna positioning with respect to site defects will randomize these effects as well.[6, 9] Thus, there is a high probability that there will be "worst case" points throughout the frequency band which will capture the site imperfection effects in a Type A analysis. However, there is always the likelihood of a significant systematic effect which is not easily determined. It should also be noted that NIS-81 classifies site imperfections as a rectangular distribution. This is largely due to the fact that existing site verification techniques do not provide for individual determination of the random and systematic contributions from the site.

For EMC measurements, a good pair of antennas may be used to determine the normalized site attenuation (NSA) of a test site and use the deviation of that value from theoretical for the site contribution. However, in the case of antenna calibration, this option is circular. Since the uncertainty of an NSA measurement can be no better than that of the antenna calibration, the uncertainty of the resulting antenna calibration could never be better than the NSA measurement plus its uncertainty!

Two options remain for antenna calibrations. The first is to attempt "round-robin" testing to compare one site to others and use the average value as the perfect site. The second method, to be published as an amendment to CISPR 16-1 in 1998, involves using calculable dipoles to verify that a site matches a perfect theoretical ground plane through an exhaustive sequence of tests.[10]

Comparisons between antenna measurements made using the same test system on the NIST (Boulder, CO) OATS and another OATS show the variation between the sites to be on the order of 0.5 dB. This variation is of the same order of magnitude as the random uncertainty contribution and thus an individual test is insufficient to draw a conclusion on the systematic error contribution of the site. It should be reiterated here that the assumption of a "golden site" for comparison purposes is not recommended. Instead the use of round-robin testing for determination of a statistically perfect site, or the new CISPR method for site verification is recommended for validation of an antenna calibration site.

Total Uncertainty

The combined standard uncertainty, $u_c(y)$, of a quantity y is computed from the square root of the sum of squares (RSS) of the individual contributions $u(x_i)$. If an individual uncertainty component does not directly correspond to the measurand, it must first be converted into the proper form by applying the appropriate conversion factor or function, $u_i(y) = c_i \bullet u(x_i)$. For example, a positioning uncertainty in centimeters must be converted into its effect on the field in dB before it can be applied to the antenna factor uncertainty. Then, for N uncorrelated individual components, the combined standard uncertainty is:

$$u_c(y) = \sqrt{\sum_{i=1}^{N} u_i(y)}$$

The expanded measurement uncertainty, U, is then determined by multiplying the combined standard uncertainty by the desired coverage factor, k, which determines the level of confidence in the uncertainty value. Thus, $U = k \bullet u_c(y)$. For the recommended 95% level of confidence, k = 2.

Using 0.5 dB for the expanded uncertainty of the random contribution and 0.75 dB as a rectangular distribution for the remaining contribution due to site imperfections and any other systematic effects,

$$u_c(y) = \sqrt{\left[\frac{0.5}{2}\right]^2 + \frac{0.75^2}{3}} = 0.5 \text{ dB}$$

Using k = 2 results in a total combined expanded measurement uncertainty of 1.0 dB. Since the random error contribution is a significant portion of the total ($u_c(y)/u(q_k)<3$), it is necessary to use an adjusted value for the coverage factor, k_p. In this case, the value is around 2.01, but in the case of only a few samples, this value could be as large as 3 to 14.

Conclusion

While there are inherent difficulties in performing a Type A analysis of a test setup, it is important not to dismiss the concept altogether. It is a relatively simple matter to obtain sufficient measurement data to produce an acceptable measure of the total random contribution to the uncertainty. This has the advantage of providing a measured value and thus limits the number of assumptions necessary to arrive at a total expanded uncertainty value. The ability to prove uncertainty claims with measured data is likely to become more important as new EMC regulations are put into effect.

References

1. NIS 81, Edition 1, *The Treatment of Uncertainty in EMC Measurements*, NAMAS, May 1994.
2. Barry N. Taylor and Chris E. Kuyatt, NIST Technical Note 1297, 1994 Edition, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST.
3. C63 EMC Measurement Uncertainty Workshop, April 29-30, 1996.
4. Manfred Stecher, "Measurement Uncertainty in EMC," *ITEM 1997*, pp. 58-64, 246, 1997.
5. Edwin L. Bronaugh and John D. M. Osburn, "A Process for the Analysis of Measurement and Determination of Measurement Uncertainty in EMC Test Procedures," 1996 International Symposium on EMC, Santa Clara, CA, August 19-23, 1996, pp. 245-249.
6. Heinrich Garn, "Uncertainties in the Calibration of Broadband EMI Test Antennas by Using a Reference Antenna," 1997 Mode-Stirred Chamber, Anechoic Chamber, and OATS Operators Meeting, April 28-May 1st, 1997.
7. Brent Dewitt, Daniel D. Hoolihan, "An Analysis of Measurement Uncertainty with Respect to IEC 1000-4-3," 1997 International Symposium on EMC, Austin, TX, August 18-22, 1997, pp. 365-369.
8. Edwin L. Bronaugh, "Uncertainty of the Frequency Spectra Produced by a Reference Radiator," Compliance Engineering, 1997 Reference Guide, pp. A101-A108.
9. Peter McNair, "Understanding Open Area Test Site Performance," 1997 Antenna Measurement Techniques Association Symposium, November 17-21, 1997, Boston, MA, pp. 25-30.
10. Jasper J. Goedbloed, "Progress in Standardization of CISPR Antenna Calibration Procedures," 1995 International Zurich Symposium on EMC, March 7-9, 1995, pp. 111-116.

***Dr. Michael D. Foegelle** received his PhD in Physics from the University of Texas at Austin, where he performed theoretical and experimental research in both condensed matter physics and EMC. He performed contract EMC research with Dr. J.D. Gavenda of UT for IBM and Rayproof where he helped to develop a semi-anechoic chamber modeling system. In 1994 he began working for EMCO (now EMC Test Systems, L.P.) in Austin Texas, where he is a physicist and principal design engineer. There, he developed the advanced calibration software used by the EMC Test Systems calibration laboratory, and a number of new products, including a high-speed isotropic field probe and a virtual device controller. He has authored or co-authored a dozen papers in the areas of EMC and condensed matter physics. (512) 835-4684, ext. 650.*