ASME B89.7.3.3-2002

GUIDELINES FOR ASSESSING THE RELIABILITY OF DIMENSIONAL MEASUREMENT UNCERTAINTY STATEMENTS

AN AMERICAN NATIONAL STANDARD

The American Society of Mechanical Engineers

The American Society of Mechanical Engineers

AN AMERICAN NATIONAL STANDARD

GUIDELINES FOR ASSESSING THE RELIABILITY OF DIMENSIONAL MEASUREMENT UNCERTAINTY STATEMENTS

ASME B89.7.3.3-2002

Date of Issuance: February 21, 2003

This Standard will be revised when the Society approves the issuance of a new edition. There will be no addenda issued to this edition.

ASME will issue written replies to inquiries concerning interpretation of technical aspects of this Standard.

This code or standard was developed under procedures accredited as meeting the criteria for American National Standards. The Standards Committee that approved the code or standard was balanced to assure that individuals from competent and concerned interests have had an opportunity to participate. The proposed code or standard was made available for public review and comment that provides an opportunity for additional public input from industry, academia, regulatory agencies, and the public-at-large.

ASME does not "approve," "rate," or "endorse" any item, construction, proprietary device, or activity.

ASME does not take any position with respect to the validity of any patent rights asserted in connection with any items mentioned in this document, and does not undertake to insure anyone utilizing a standard against liability for infringement of any applicable letters patent, nor assume any such liability. Users of a code or standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, is entirely their own responsibility.

Participation by federal agency representative(s) or person(s) affiliated with industry is not to be interpreted as government or industry endorsement of this code or standard.

ASME accepts responsibility for only those interpretations of this document issued in accordance with the established ASME procedures and policies, which precludes the issuance of interpretations by individuals.

The American Society of Mechanical Engineers
Three Park Avenue, New York, NY 10016-5990

CONTENTS

FOREWORD

The ISO Guide to the Expression of Uncertainty in Measurement (GUM) is now the internationally-accepted method of expressing measurement uncertainty. The U.S. has adopted the GUM as a national standard. (See ANSI/NCSL Z540-2.) The evaluation of measurement uncertainty has been applied for some time at national measurement institutes but more recently issues such as measurement traceability and laboratory accreditation are resulting in its widespread use in calibration laboratories.

Given the potential impact to business practices, national and international standards committees are working to publish new standards and technical reports that will facilitate the integration of the GUM approach and the consideration of measurement uncertainty. In support of this effort, ASME B89 Committee for Dimensional Metrology has formed Division 7, Measurement Uncertainty.

Measurement uncertainty has important economic consequences for calibration and measurement activities. In calibration reports, the magnitude of the uncertainty is often taken as an indication of the quality of the laboratory, and smaller uncertainty values generally are of higher value and of higher cost. In the sorting of artifacts into classes or grades, uncertainty has an economic impact through the use of decision rules. ASME B89.7.3.1, Guidelines to Decision Rules in Determining Conformance to Specifications, addresses the role of measurement uncertainty when accepting or rejecting products based on a measurement result and a product specification.

With increasing use of measurements from laboratories that are accredited, and subsequent measurement uncertainty statements, significant economic interests are at stake, so it is not surprising that metrologists might disagree over the magnitude of the measurement uncertainty statements. While the selection of a decision rule is a business decision, the evaluation of the measurement uncertainty is a technical activity. This report provides guidance for resolving disagreements involving measurement uncertainty statements.

This report was approved by the American National Standards Institute on April 22, 2002.

Comments and suggestions for improvement of this Technical Report are welcomed. They should be addressed to: ASME, Secretary, B89 Committee, Three Park Avenue, New York, NY 10016-5990

ASME STANDARDS COMMITTEE B89
Dimensional Metrology

(The following is the roster of the Committee at the time of approval of this Standard.)

OFFICERS

B. Parry, *Chair*
D. Beutel *Vice Chair*
M. Lo, *Secretary*

COMMITTEE PERSONNEL

J. B. Bryan, Bryan Associates
T. Carpenter, US Air Force Metrology Labs
T. Charlton, Jr., Charlton Associates
G. A. Hetland, International Institute of GDNT
R. J. Hocken, University of North Carolina
M. Liebers, Professional Instruments
S. D. Phillips, NIST
J. Salsbury, Mitutoyo America
D. A. Swyt, NIST
B. R. Taylor, Renishaw PLC
R. C. Veale, NIST

SUBCOMMITTEE 7: MEASUREMENT UNCERTAINTY

G. Hetland, *Chair,* International Institute of GDNT
D. Swyt, *Vice Chair,* NIST
W. Beckwith, Brown & Sharpe
D. Beutel, Caterpillar
J. Buttress, Hutchinson Technology, Inc.
T. Charlton, Charlton Associates
W. T. Estler, NIST
H. Harary, NIST
B. Parry, The Boeing Company
S. Phillips, NIST
J. Raja, University of North Carolina
P. Stein, P. G. Stein Consultants

PROJECT TEAM ON DECISION RULES

S. Phillips, *Chair,* NIST
J. Buttress, Hutchinson Technology
T. Carpenter, US Air Force Metrology Labs
W. T. Estler, NIST
H. Harary, NIST
B. Parry, The Boeing Company

CORRESPONDENCE WITH B89 COMMITTEE

General. ASME Codes and Standards are developed and maintained with the intent to represent the consensus of concerned interests. As such, users of this Standard may interact with the Committee by requesting interpretations, proposing revisions, and attending Committee meetings. Correspondence should be addressed to:

Secretary, B89 Main Committee
The American Society of Mechanical Engineers
Three Park Avenue
New York, NY 10016

Proposing Revisions. Revisions are made periodically to the Standard to incorporate changes that appear necessary or desirable, as demonstrated by the experience gained from the application of the Standard. Approved revisions will be published periodically.

The Committee welcomes proposals for revisions to this Standard. Such proposals should be as specific as possible, citing the paragraph number(s), the proposed wording, and a detailed description of the reasons for the proposal, including any pertinent documentation.

Interpretations. Upon request, the B89 Committee will render an interpretation of any requirement of the Standard. Interpretations can only be rendered in response to a written request sent to the Secretary of the B89 Main Committee.

The request for interpretation should be clear and unambiguous. It is further recommended that the inquirer submit his/her request in the following format:

Subject:	Cite the applicable paragraph number(s) and the topic of the inquiry.
Edition:	Cite the applicable edition of the Standard for which the interpretation is being requested.
Question:	Phrase the question as a request for an interpretation of a specific requirement suitable for general understanding and use, not as a request for an approval of a proprietary design or situation. The inquirer may also include any plans or drawings which are necessary to explain the question; however, they should not contain proprietary names or information.

Requests that are not in this format will be rewritten in this format by the Committee prior to being answered, which may inadvertently change the intent of the original request.

ASME procedures provide for reconsideration of any interpretation when or if additional information that might affect an interpretation is available. Further, persons aggrieved by an interpretation may appeal to the cognizant ASME Committee or Subcommittee. ASME does not "approve," "certify," "rate," or "endorse" any item, construction, proprietary device, or activity.

Attending Committee Meetings. The B89 Main Committee regularly holds meetings, which are open to the public. Persons wishing to attend any meeting should contact the Secretary of the B89 Main Committee.

GUIDELINES FOR ASSESSING THE RELIABILITY OF DIMENSIONAL MEASUREMENT UNCERTAINTY STATEMENTS

ABSTRACT

The primary purpose of this technical report is to provide guidelines for assessing the reliability of measurement uncertainty statements. Applying these guidelines can assist businesses in avoiding disagreements about measurement uncertainty statements and in resolving such disagreements should they occur. Disagreements over uncertainty statements involving both single measurement systems and multiple measurement systems (each having their own uncertainty statement) are considered. Guidance is provided for examining uncertainty budgets as the primary method of assessing their reliability. Additionally, resolution by direct measurement of the measurand is also discussed.

1 SCOPE

1.1 Objective

This technical report provides guidance in assessing the reliability of a statement of measurement uncertainty in question, that is, in judging whether that stated uncertainty can be trusted to include the values that could reasonably be attributed to the measured quantity (measurand) with which that stated uncertainty is associated.

1.2 Applicability

This report is most applicable to statements of uncertainty in the results of dimensional measurements based upon the ISO Guide to Expression of Uncertainty in Measurement (GUM). (Also called ANSI/NCSL Z540-2.)

1.3 Purpose

This technical report helps parties to avoid potential, or resolve actual, disagreements over the magnitude of a stated measurement uncertainty, particularly when that uncertainty is part of a determination of conformity of a manufactured product to a dimensional specification.

2 DEFINITIONS[1]

acceptance zone: the set of values of a characteristic, for a specified measurement process and decision rule, that results in product acceptance when a measurement result is within this zone.[2]

decision rule: a documented rule, meeting the requirements of section 3 of ASME B89.7.3.1, that describes how measurement uncertainty will be allocated with regard to accepting or rejecting a product according to its specification and the result of a measurement.

expanded uncertainty: quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand. See GUM, 2.3.5.

guard band: the magnitude of the offset from the specification limit to the acceptance or rejection zone boundary.[3, 4, 5, 6, 7, 8]

[1] Many of these definitions are selected from ASME B89.7.3.1. The figures from that document are omitted here for brevity.

[2] When claiming product acceptance, it is important to state the decision rule; e.g., "acceptance using the XX rule."

[3] The symbol g is deliberately used for the guard band, instead of the symbol U employed in ISO 14253-1 since U is reserved for the expanded uncertainty which is associated with a measurement result and hence it is confusing to attach U to a specification limit. The evaluation of U is a technical issue, while the evaluation of g is a business decision.

[4] The guard band is usually expressed as a percentage of the expanded uncertainty, i.e., a 100% guard band has the magnitude of the expanded uncertainty U.

[5] Two-sided guard banding occurs when a guard band is applied to both the upper and lower specification limits. (In some exceptional situations the guard band applied within the specification zone, g_{In}, could be different at the upper specification limit and at the lower specification limit. This would reflect a different risk assessment associated with an upper or lower out-of-specification condition depending on whether the characteristic was larger or smaller than allowed by the specification zone.) If both the upper and lower guard bands are the same size then this is called symmetric two-sided guard banding.

[6] A guard band is sometimes distinguished as the upper or lower guard band, associated with the upper or lower specification limit. Subscripts are sometimes attached to the guard band notation, g, to provide clarity, e.g., g_{Up} and g_{Lo}. See ASME B89.7.3.1, Fig. 1.

measurand: particular quantity subject to measurement. See VIM 2.6.[9]

N:1 decision rule: a situation where the width of the specification zone is at least N times larger than the uncertainty interval for the measurement result.[10]

rejection zone: the set of values of a characteristic, for a specified measurement process and decision rule, that results in product rejection when a measurement result is within this zone.[11]

specification zone (of an instrument or workpiece): the set of values of a characteristic between, and including, the specification limits.[12, 13, 14]

stringent acceptance: the situation when the acceptance zone is reduced from the specification zone by a guard band(s).[15, 16]

stringent rejection: the situation when the rejection zone is increased beyond the specification zone by a guard band.[17]

uncertainty interval (of a measurement): the set of values of a characteristic about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.[18, 19]

[7] The guard band, g, is always a positive quantity; its location, e.g., inside or outside the specification zone, is determined by the type of acceptance or rejection desired. See ASME B89.7.3.1, Section 4.

[8] While these guidelines emphasize the use of guard bands, an equivalent methodology is to use gauging limits as in ASME B89.7.2-1999.

[9] The specification of a measurand may require statements about such quantities as time, temperature, and pressure.

[10] A common example is the 4:1 ratio.

[11] When claiming product rejection, it is important to state the decision rule; e.g., "rejection using the XX rule."

[12] The width of the specification zone is a positive number.

[13] In the case of workpieces, the width of the specification zone is identical to the tolerance.

[14] Specification zone is equivalent to "tolerance interval" or "tolerance zone" defined in ISO 3534-2.

[15] Stringent acceptance and relaxed rejection occur together in a binary decision rule.

[16] The stringent acceptance zone is analogous to the conformance zone described in ISO 14253-1.

[17] Relaxed acceptance and stringent rejection occur together in a binary decision rule.

[18] The width of the uncertainty interval is typically twice the expanded uncertainty.

[19] The uncertainty interval for the mean of repeated measurements may decrease with increasing numbers of measurements.

3 THE NATURE OF DISAGREEMENTS IN UNCERTAINTY STATEMENTS

3.1 General

In an ideal situation, customers and suppliers will address the issue of measurement uncertainty when they discuss the product specifications. Agreeing on the measurement plan, the corresponding magnitude of the measurement uncertainty, and the decision rule (if applicable), will avoid future disagreements regarding the acceptance/rejection of a product. However, it is recognized that two experts can produce two different uncertainty statements often varying as much as 25% due to differing assumptions and data (as described in section 5). Resolving these differences at the contract stage is potentially less contentious than doing so after an argument develops over the acceptance or rejection of the product.

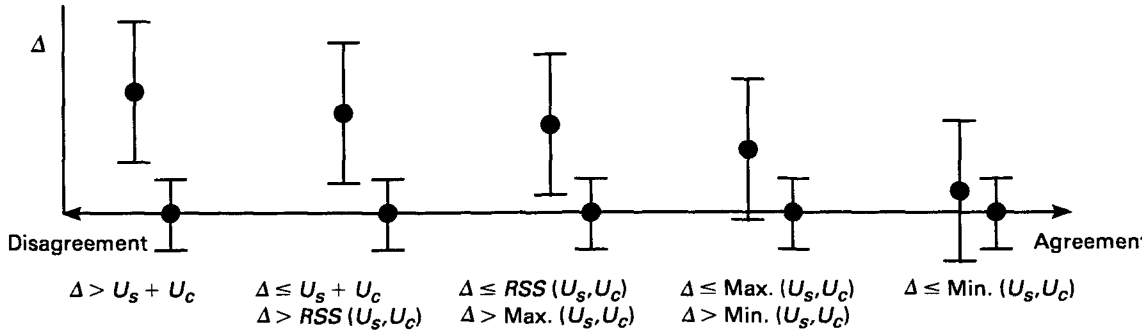
3.2 Disagreements Involving Single Measurement Systems

In many situations there is only a single measurement system; e.g., a customer agrees to accept the supplier's measurement results provided that the supplier uses stringent acceptance with a 100% guard band (i.e., the guard band equals the expanded uncertainty). In this example, a disagreement may arise if the customer feels the supplier has underestimated the measurement uncertainty. Although there is a single measurement system, the supplier and the customer have developed differing uncertainty statements.

3.3 Disagreements Involving Multiple Measurement Systems

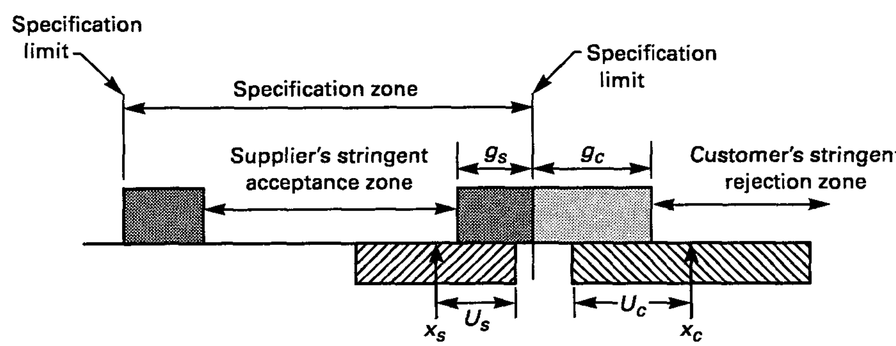
In some situations, a customer and supplier both make measurements, each having their own measurement system and uncertainty statement. There are two cases to consider: first, when a product characteristic is being measured to assign it a value, e.g., the length of a gauge block, and second, when a product characteristic is being measured to determine whether it conforms with specifications.

In the first case, a best estimate of the value of the product characteristic is being sought. Two measurements, from different measurement systems, will give a better estimate when their results are appropriately combined than will each system independently, provided the uncertainty statements associated with the measurement systems are valid. It is unlikely that the measurements performed by the supplier and the customer



GENERAL NOTE: Five pairs of measurement, one from a customer and one from a supplier, illustrating different degrees of measurement agreement and disagreement; the uncertainty bars represent the expanded uncertainty of the associated measurements.

FIG. 1 EXAMPLES OF MEASUREMENT AGREEMENT AND DISAGREEMENT



GENERAL NOTE: An example of disagreement over the conformance of a product. The supplier, using stringent acceptance, has measurement result x_s and claims the product to be acceptable; the customer, using a different measurement system and a stringent rejection decision rule has measurement result x_c and claims product rejection.

FIG. 2 EXAMPLE OF PRODUCT CONFORMANCE DISAGREEMENT

will yield exactly the same value; however, agreement between the measurements is obtained by some extent of overlap of the uncertainty intervals. The extent of overlap should be specified in order to clearly identify when the parties are in disagreement. (This avoids disagreements on what constitutes a measurement disagreement.) There are several possible cases of metrological significance as shown in Fig. 1. Let x_s and x_c be the measurement results of the supplier and customer, with respective expanded uncertainties of U_s and U_c (both using a coverage factor of $k = 2$). Let $\Delta = |x_s - x_c|$ be the absolute value of the difference between the measurements. Figure 1 illustrates this case with five different pairs of measurements. The measurements are considered to be in disagreement when $\Delta > U_s + U_c$ and in agreement when Δ is less than the minimum of either U_s or U_c. In laboratory round robins, measurements are generally considered to agree when Δ is less than or equal to the root sum of squares (RSS) of the two expanded uncertainties and in disagreement if Δ is greater than this quantity.

Sometimes two different measurement systems are used to determine if a product is in conformance with specifications and the outcomes of the two measurements differ, i.e., the acceptance or rejection of the product is in dispute. This case can be separated into issues involving the decision rule and issues involving the reliability of the uncertainty statement. When two different parties each perform measurements on the same product, potential disagreements can arise due to inherent conflict in the decision rules. For example, if both the supplier and customer apply stringent acceptance then the party with the larger guard band will typically reject more of the product. This concerns the decision rule selection, not the uncertainty statements, and consequently is outside the scope of these guidelines.

Alternatively, reasonable decision rules may be selected, e.g., the supplier is using stringent acceptance and the customer is using stringent rejection and yet the measurement outcomes differ. For example, the supplier claims the product is acceptable and the customer claims the product is rejectable. One potential source of this dispute is that one or both parties have incorrectly computed the guard band due to an incorrect assessment of their measurement uncertainty. Figure 2 illustrates the supplier claiming conformance using stringent acceptance with a 100% guard band ($g_s = U_s$) and the customer claims nonconformance using stringent rejection with a 100% guard band ($g_c = U_c$). In this scenario it is likely that one or both of the uncertainty statements are invalid.

4 CAUSES OF DISAGREEMENT IN MEASUREMENT RESULTS HAVING UNCERTAINTY STATEMENTS

4.1 General

Although the customer and supplier may disagree over the measurement uncertainty during the contract negotiations, the more common and contentious case involves measurements performed by the customer and supplier where the results are in disagreement. The customer may reject the product and the supplier may claim it is acceptable. There are several possible explanations for this disagreement.

4.2 Blunders

One simple explanation of how two measurement results could be in disagreement is that at least one of the results includes a measurement blunder. Typical examples include typographical errors in recording or transcribing the measurement result, improper instrument settings, and failure to properly fixture the product under test. The possibility of committing a blunder is not to be included in the uncertainty budget (GUM 3.4.7). If two measurement results differ by a large amount (relative to the RSS of the expanded uncertainties) a blunder is suspected and should be investigated. This may include examining the original measurement records, instrument settings, operator, fixturing, other sources of gross error, or simply repeating the measurements. A blunder typically results in a measurement outlier. This topic is briefly discussed in ASME B89.7.3.1, Appendix C. Since these blunders are not associated with the development of an uncertainty budget they will not be further considered in this document.

4.3 GUM Noncompliance and Uncorrected Systematic Errors

The GUM provides a unified, consistent means for calculating measurement uncertainty. Failure to follow the procedures described in the GUM may result in a substantially different uncertainty statement. Procedures for calculating measurement uncertainty are well described in the GUM, a nationally and internationally recognized document. Deviations from this approach, e.g., algebraically summing uncertainty components, will lead to disagreements between uncertainty budgets and should be avoided. Common ways of failing to comply with the GUM include not accounting for known systematic errors (GUM 3.2 and F.2.4.5) and not using the law of propagation of uncertainty (GUM 5.1.2) (or some other appropriate means of combining uncertainty sources such as computer simulation). While it is recommended that corrections for all known systematic errors be applied to the measurement result, in some cases this is economically undesirable; a discussion of the inclusion of uncorrected systematic errors in uncertainty statements can be found in Phillips, Eberhardt, and Parry.

4.4 Poorly Realized or Incompletely Defined Measurand

Measurement results and their uncertainties are associated with a particular measurand. The numerical value associated with some measurands may be time dependent; i.e., the value realized by a measurement could change in time due to degradation, temporal instability, wear, or damage. Two measurements separated in time could realize two different values for a well-specified measurand. A common example is damage to the product under consideration (particularly in transport) which systematically changes the value associated with the measurand. (The measurand itself, which is a set of specifications, remains unchanged; see Phillips, Estler, et al. for a more extensive discussion of this issue.) One method of detecting this problem involves examining the consistency of repeated measurements separated in time, e.g., before and after transport.

In order for measurement results and their associated measurement uncertainties to be compared, they must be measuring the same quantity (the same measurand). A measurand should be defined with sufficient completeness with respect to the required accuracy, so that for all practical purposes associated with the measurements its value is unique (GUM 3.1.1). If the measurand is poorly defined, different measurement methods may produce different measurement results, all of which are consistent with the incompletely defined measurand. This is sometimes called the "methods divergence" problem. The uncertainty associated with an incompletely defined measurand is to be included in the uncertainty budget. The ambiguity in the definition of the measurand and the stability of its realization should be assessed before an uncertainty budget can be created or compared with another.

4.5 Statistically Rare Measurement Results

Although the uncertainty interval will contain a large fraction of the values that could reasonably be attributed to the measurand, it does not contain all such values. Using a coverage factor of two ($k = 2$) will yield, on a statistical basis, approximately five cases per 100 where two measurement will differ by more than the RSS of the expanded uncertainties. Similarly, there will be approximately five cases per 1,000 where two measurements will differ by more than the sum of the expanded uncertainties. In general, the difference of these statistically rare measurements is unlikely to be large compared to the RSS of the expanded uncertainties. Repeated measurements should resolve this issue, as they are likely to yield mean values that differ by less than the RSS of the expanded uncertainties, unless the uncertainties are dominated by systematic errors.

4.6 Incomplete Uncertainty Statements

If an uncertainty budget has failed to account properly for all significant uncertainty sources, then the interval defined by the expanded uncertainty will not encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand. In this case, the two measurement results could be in disagreement, depending upon the extent of missing uncertainty sources in the budgets.

5 METHODS OF RESOLUTION

5.1 General

This section provides guidance on resolving disagreements over the uncertainty statements developed by the customer and the supplier. First, the significance of the disagreement is to be established. The two primary means of resolution are: (1) comparison of uncertainty budgets, which involves a discussion of the uncertainty sources, their magnitudes, and their effects on the measurement result, and (2) resolution by direct measurement of the measurand (if possible).

Generally, comparison of the uncertainty budgets is the most expedient route to resolving disagreements. Resolution through direct measurement of the measurand typically invokes the definition of the uncertainty statement, i.e., a large fraction of the measurement errors are contained within the expanded uncertainty interval. These guidelines recognize that demonstrating the invalidity of an uncertainty statement is easier than demonstrating its validity.

5.2 Significance of Disagreement

In all cases of disagreement, the issue of significance should be considered first. Due to the multitude of uncertainty sources, and the assumptions involved in assessing their magnitudes and impacts on the measurement result, it is common for two experts to arrive at two different uncertainty statements. Differences in the stated uncertainty value on the order of 25% are not unusual and reflect slightly different assumptions present in the uncertainty budget. For example, variations in the Type A standard uncertainty based on ten observations can vary easily by 25% (GUM E.4). Similarly, Type B uncertainties can vary by 25% depending on the type of distribution assumed or the knowledge used in the estimated extent of the distribution. Typically, differences at this level often are deemed insignificant. Unless significant evidence demonstrates that one set of assumptions is more relevant than another, it is recommended that uncertainty budgets differing by 25% or less use the mean of the two uncertainty values.

For larger discrepancies between uncertainty statements, this guideline recommends that each competing uncertainty budget be expressed as a percentage of the applicable specification. While differences between uncertainty statements on the order of 25%–50% may occur, these differences might not be significant when each is considered as a percentage of the specification zone, and hence their effect on the acceptance or rejection of products. For example, if one uncertainty statement results in a 10:1 decision rule and another uncertainty budget results in a 15:1 decision rule, the amount of product affected by these small guard bands may be insignificant. If so, then unless significant evidence can be presented to demonstrate the superiority of one uncertainty budget over the other, it is recommended to use the mean of the two uncertainty values. If the difference is still deemed significant, then a comparison of uncertainty budgets should be initiated.

5.3 Comparison of Uncertainty Budgets

Comparing uncertainty budgets is generally the first step in resolving discrepancies that are deemed significant. Such comparisons consist of verifying that the uncertainty budget includes, and properly accounts for, all significant sources of uncertainty.

5.3.1 Accounting for Uncertainty Sources

5.3.1.1 Influence Quantities. The factors that affect measurement results are known as influence quantities (GUM 3.1. and GUM B.2.10). All significant influence quantities should be listed and a comparison made to ensure that both uncertainty budgets have considered these effects. If an influence quantity is accounted for in one budget but omitted in the other, discussion regarding the significance of the quantity should be conducted. Unless the uncertainty budget is primarily composed of numerous small uncertainty sources, individual sources that have a standard uncertainty less than 10% of the largest standard uncertainty source in the budget can usually be omitted without significance.

The list of influence quantities also depends on the time period over which the uncertainty budget is valid. For example, if an uncertainty budget is developed for a single measurement then only those influence quantities that affect that measurement need be considered. However, if an uncertainty budget is to be used for many measurements (of the same measurand) then all influence quantities over the entire period when measurements will be taken must be considered. For example, seasonal temperature changes might be a relevant uncertainty source for an uncertainty budget intended to be valid for product measurements made over the course of a year.

5.3.1.2 Input Quantities. Various uncertainty sources known as input quantities comprise an uncertainty budget. These quantities are listed and combined using the methods described in the GUM. Each input quantity is composed of one or more influence quantities. Metrologists choose how different influence quantities will be combined into input quantities and this is why two uncertainty budgets, both of which are correctly constructed, can appear to be very different in their treatment of the influence quantities. For example, one budget might list each influence quantity as an input quantity and have a long list of uncertainty sources that need to be combined. Another budget might have only a few input quantities, choosing to combine many influence quantities into a single input quantity, such as a Type A uncertainty evaluated from a long-term reproducibility study. The important point is that all influence quantities of the measurement are accounted for in some input quantity. In certain cases, e.g., uncertainty evaluated by computer simulation, input quantities are represented as parameters that are allowed to vary between simulation cycles.

5.3.2 Magnitudes of Uncertainty Components (Standard Uncertainties of Input Quantities).

After ascertaining that each significant influence quantity is present in the uncertainty budget in some input quantity, the input quantities must be evaluated. The magnitudes of uncertainty components are quantified by standard deviations, known as standard uncertainties (GUM 4.1). Each input quantity has an associated standard uncertainty. The value of the standard uncertainty must account for the range of values that could reasonably be attributed to the input quantity over the time period that the measurements are performed. Some uncertainty budgets pertain to a single measurement that occurs in a short period of time. Other uncertainty budgets might pertain to many subsequent measurements, (as is typical of production workpieces), where the measurement conditions, while being bounded, change from measurement to measurement.

The time scale over which the measurements are performed must be considered. An uncertainty budget designed for a large number of measurements, e.g., production workpieces that are continuously produced and inspected over a long time scale (days, weeks, months, or years), will have greater variation in the input quantities than measurements performed in a short time period. Consequently, when determining the standard uncertainties of the input quantities it is essential to consider the full range of possible variations that may occur during the measurements. In addition, for uncertainty budgets that will apply to many subsequent measurements, the bounds on the permissible variation of the input quantities under which the uncertainty budget is valid should be clearly stated.

Type A standard uncertainties are evaluated using statistical means that are generally well defined; however, it is crucial that the data fully represent the input quantity to be quantified. For example, if the uncertainty source associated with measurement reproducibility depends upon the operator, then several operators (not just one) must be included in the reproducibility (Type A) data. Differences in Type A standard uncertainties usually result from failing to allow the input quantity to vary over the entire range of values permitted in the uncertainty budget. For practical reasons it may not be possible to vary all input quantities over their full extent; for such a truncated Type A reproducibility study, Type B standard uncertainties may be used to account for any additional unobserved variation and will appear as an additional input quantity in the uncertainty budget.

For Type B standard uncertainties, a distribution must be assumed in order to obtain a standard uncertainty. As with Type A uncertainties, the assumed distribution must characterize the range of values that could reasonably be attributed to the uncertainty source. Disagreement over Type B uncertainties can sometimes be resolved by consulting reference books, technical papers and reports, or other documented material regarding the typical range of values associated with the input quantity. Unless additional information is known, a Type B standard uncertainty usually is assigned by assuming a normal or uniform distribution (GUM 4.3).

5.3.3 Effects of Uncertainty Sources (Sensitivity Coefficients). Once an input quantity has been deemed to be a significant uncertainty source and its standard uncertainty has been determined, its impact on the result of the measurement must be estimated. The sensitivity of the same measurand to the same uncertainty source may vary widely, depending on details of the measurement. For example, the uncertainty in the length of a block of material due to temperature measurement uncertainty is more than an order of magnitude larger for a block of plastic than for one of ceramic. A discussion of several methods for determining sensitivity coefficients and difficulties with the methods follows.

One method of obtaining sensitivity and correlation coefficients is by taking partial derivatives of an analytic mathematical model of the measurement process. Discrepancies between two sensitivity coefficients using this method are due to different mathematical models. Therefore, examination of the accuracy and completeness of the mathematical model and the reasonableness of its behavior (as the input quantities are varied) is recommended. Particular attention should be paid to issues involving correlation between uncertainty sources as this is a factor often omitted in mathematical models.

Many uncertainty sources do not lend themselves to analytic mathematical models. For example, the effect of loose fixturing may not have a simple mathematical formula, and consequently without a mathematical model partial derivatives cannot be taken. In these cases the metrologist often uses, in effect, a numerical evaluation of the derivative by varying the input quantity by some small known amount and observing the change in the measurement result. The ratio of measurement change to input quantity change is the sensitivity coefficient. Care must be taken to determine this value accurately. For example, the induced and observed changes must be significantly larger than their associated uncertainties. An effective version of this method involves varying the input quantity over the full range of values permitted in the uncertainty budget, in a manner consistent with the uncertainty source, i.e., using the same distribution of values. In this case, the standard deviation of the measurement results is the product of the sensitivity coefficient with the standard uncertainty of that source and includes some correlation effects.

Another method of assessing the impact of an uncertainty source is through computer simulation. Simulation typically involves a mathematical model of the measurement process expressing the output quantity (measurand) as a function of the input quantities (GUM 4.1.1). Instead of taking partial derivatives to calculate the sensitivity coefficients, the output is repeatedly calculated for different combinations of input quantities. The input quantities usually are represented by their distribution of values, hence a comprehensive simulation will sample over the full range and possible combinations of input quantities (given by their distributions) resulting in a large number of (slightly different) output quantities. The standard deviation of these output quantities represents the standard uncertainty of the measurand with regard to the input quantity uncertainty sources. This technique can capture complex correlations between uncertainty sources that might otherwise be difficult to calculate.

Disagreement over an uncertainty statement produced by a computer simulation can arise for several reasons.

(a) The simulation accounts only for sources of uncertainty that are included in the mathematical model of the measurement process. For most actual measurements, sources of uncertainty which are not easily modeled are also present and must be included in the uncertainty statement by alternative means. Therefore the comprehensiveness (accounting for all uncertainty sources) of both the mathematical model and the entire uncertainty statement should be checked.

(b) Two metrologists may model a measurement in very different ways depending on the information available, for example what parameters are considered input quantities.

While detailed mathematical verification of the models is usually too complex for most practitioners, the reasonableness of the simulations often can be established by simulating measurements with known results.

For example, if it is known that the measurement of calibrated artifacts always produces results within certain limits, e.g., manufacturer's specifications, then the simulation of this process should result in a predicted uncertainty that contains these limits. If the uncertainty from the simulation (after inclusion of other relevant uncertainty sources) does not include the known errors of this (special case) measurement, then the computed uncertainty is too small and the simulation model must be reexamined.

5.3.4 Third Party Review and Accreditation. If a disagreement persists after a review of the uncertainty budgets, a third party review of the budgets may be requested. A third party review may bring specific expertise relevant to the uncertainty budget, and may identify problems that have escaped the attention of previous reviews. Additionally, a third party brings an unbiased opinion to the review process.

Some types of laboratory accreditation require the examination of uncertainty budgets and the demonstration of measurement competency, representing a form of third party review. If accreditation is in the field of the measurement under consideration and the uncertainty claimed is no smaller than the accreditation documentation, then this can be used as evidence to support an uncertainty statement. The burden is on the other party to demonstrate that the measurement under consideration is sufficiently different, so that the accreditation is not applicable.

Accreditation uncertainties usually represent "best practice," defined as the smallest uncertainty that a laboratory can achieve when performing routine calibrations of nearly ideal artifacts. If it can be shown that the measurements were not conducted using "best practices," e.g., the part under consideration deviates significantly from ideal form, then the weight of accreditation as evidence for the uncertainty budget is diminished.

5.4 Direct Measurement of the Measurand

5.4.1 General. In addition to comparing uncertainty budgets, discrepancies may be resolved by direct measurement of the measurand. (For some types of measurands, such as destructive testing, this may not be possible.) In contrast to measuring input quantities, e.g., as standard uncertainties, this procedure appeals directly to the definition of an uncertainty statement. In general, measurements are time consuming and costly, so this procedure is recommended only after a "paper" comparison of uncertainty budgets has been conducted. In some cases, previous measurements may be relevant to the current uncertainty budget and they should be considered first because they are available and less costly. As a last resort, measurements of calibrated parts or artifacts could be performed, with the associated errors examined with regard to the uncertainty statement. If a significant fraction of the errors lie outside the expanded uncertainty interval the measurement uncertainty statement is invalid. As stated earlier, it is easier to demonstrate the invalidity of an uncertainty statement than to demonstrate the validity.

5.4.2 Historical Measurements. In some measurement situations, large numbers of previous measurements may have been performed and recorded. The distribution of these measurement results can be used as evidence to support an uncertainty statement. Although this distribution is a convolution of measurement uncertainty and product production variation, this could represent an upper bound on the measurement uncertainty. Provided that these measurement results represent a variation over all influence quantities (discussion follows), it represents an upper bound on the uncertainty and can be used to refute a claim that the uncertainty must be larger than this value. In order for the historical measurements to represent the measurement uncertainty, the number of measurements and the period of time during which they were performed need to be sufficiently large so that all sources of uncertainty are varied over their full extent.

Specifically, measurements performed in a relatively short period of time are unlikely to show the variation due to caused by long term sources of variation, e.g., seasonal effects, and consequently do not fully represent the uncertainty. Some uncertainty components may be difficult to vary such as the thermal expansion coefficient, which is a property of the workpiece material. For these uncertainty components a Type B analysis must be performed and included in the uncertainty budget.

Additionally, the expected error of the historical measurements must be determined. The collection of historical measurements might possess systematic error. This can be determined, for example, by the periodic measurement of calibrated check standards. Any statistically significant systematic error must be treated appropriately, either by correction (preferred) or inclusion in the uncertainty statement (see Phillips, Eberhardt, and Parry). Note that the uncertainty in the systematic error is always assessed and included in the uncertainty budget.

5.4.3 Round Robins. Round robins represent a specific form of historical measurements. While the number of measurements performed by a single facility in the round robin may be small, the results can be compared to measurements made at other facilities. A round robin can be thought of as a reproducibility study where the collection of different operators, environmental conditions, and other sources of variation are included as different measurement results. This amount of variation might otherwise take years to observe in a reproducibility study conducted at a single facility.

In some round robins a conventional true value may be assigned if, for example, a National Measurement Institute (NMI) has measured the artifact and the other participants have much larger uncertainty statements than the NMI. In this case, systematic measurement error (in addition to reproducibility) also may be detected. If the artifact used in the round robin is representative of the product under consideration, and the mean or conventional true value of the round robin is within the expanded uncertainty interval of the measurement, this provides a powerful argument for the validity of the measurement and uncertainty budget.

5.4.4 Reproducibility Measurements. If new measurements are needed to resolve a disagreement over the uncertainty budget, then a reproducibility study may be the easiest step. In this study, all influence quantities that can be varied are allowed to vary over their permitted limits, while measuring the same workpiece. Typically, this includes fixturing, operators, and other nominal sources of uncertainty. The distribution of measurement results from a reproducibility study places a lower bound on the measurement uncertainty. The actual error distribution may be considerably larger because it might not be possible to vary all the input quantities. Also, there may be a systematic error that is undetected because a calibrated artifact (representing the "true value") is not used. See para. 5.4.2.

A reproducibility study involving a large number of measurements of the same artifact or workpiece should result with two standard deviations of the measurements being less than or equal to the expanded uncertainty (with $k = 2$). This is a necessary but insufficient condition of a valid uncertainty statement. This is not sufficient to prove an uncertainty statement because measurements that have systematic errors may result in a small range of values (thus a small standard deviation). In this case, two standard deviations of the measurement results will be less than the expanded uncertainty while the measurement errors could be considerably larger than the expanded uncertainty.

In some situations, a large number of repeated measurements on the same artifact or workpiece may not be available but rather two measurements on each of a large number of similar artifacts or workpieces might be known. For example, a laboratory may have measured 100 nominally identical gauge blocks twice each, and seeks to determine if the results are consistent with their uncertainty statement. In this example, let Δ_i be the difference between the two measurements on the i^{th} artifact. Let N be the total number of nominally similar artifacts each having the same expanded uncertainty. A necessary condition for a valid uncertainty statement is shown in the first half of the equation, whereas if each artifact has a different expanded uncertainty, U_i, then the necessary condition is shown in the second half. Again, this is a necessary but insufficient condition to demonstrate the validity of the uncertainty statement. Each of the two measurements comprising Δ may have the same systematic error leading to a series of small Δ's but the measurement error could be significantly larger than U.

$$2\sqrt{\frac{1}{N}\sum_{i=1}^{N}\Delta_i^2} \le \sqrt{2}U \quad \text{or} \quad \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{\Delta_i^2}{U_i^2}} \le \sqrt{2}$$

5.4.5 Measurement of Calibrated Artifacts. One of the most powerful methods to invalidate an uncertainty budget is the measurement of well-calibrated artifacts that are similar to the product under consideration. *Well-calibrated* means that the uncertainty of the artifact is small relative to that of the claimed uncertainty budget. *Similar to the product under consideration* means that all significant factors which influence the measurement result, e.g., workpiece form error, are similarly represented on the calibrated artifact. This allows for the estimation of measurement errors. Repeated measurements of calibrated artifacts should contain a large fraction (typically 95% when using a coverage factor of two) of the measurement errors within an interval equal to the RSS of the claimed expanded uncertainty and the artifact expanded uncertainty. If a significant fraction of errors lie outside this interval the claimed uncertainty budget is invalidated.

5.4.6 Third Party Measurements. An alternative to measuring a calibrated artifact is to employ a third party measurement of the product under consideration. To be useful, the uncertainty of this result should be relatively small compared to the uncertainty statements under dispute. The method is similar to measuring a calibrated artifact since the accurate third party measurement, in effect, calibrates the product under consideration. Again the difference between the measurement result and the calibrated, (third party), value is an estimate of the measurement error. This error should fall within the interval defined by the RSS of the expanded uncertainties of the disputed uncertainty statement and that of the third party measurement. Ideally many such comparisons should be conducted on a variety of parts to provide a statistical basis for confirming or invalidating an uncertainty statement. However, even with only one comparison, if the observed error falls well outside the RSS of the two expanded uncertainty statements this is strong evidence that at least one of the claimed uncertainty statements is invalid.

6 REFERENCES

ASME B89.7.2-1999, Dimensional Measurement Planning.

ASME B89.7.3.1, Guidelines to Decision Rules: Considering Measurement Uncertainty in Determining Conformance to Specifications.

Publisher: The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016-5990

ANSI/NCSL Z540-2 U.S. Guide to the Expression of Uncertainty in Measurement, National Conference of Standards Laboratories, 1997.

Publisher: The American Society for Testing and Materials (ASTM), 100 Barr Harbor Drive, West Conshohocken, PA 19428

ISO, Guide to the Expression of Uncertainty in Measurement (GUM), 1993 (corrected and reprinted 1995)

ISO, International Vocabulary of Basic and General Terms in Metrology (VIM) (Second Ed.). See 6.10 (ISBN 92-67-01075-1).

ISO 3534-2: 1993, Statistice—Vocabulary and Symbols—Part 2: Statistical Quality Control.

ISO 14253-1: 1998, Geometrical Product Specifications (GPS)—Inspection by Measurement of Workpieces and Measuring Instruments—Part 1: Decision Rules for Proving Conformance or Nonconformance with Specification.

Publisher: International Organization for Standardization (ISO), 1 rue de Varembé, Case postale 56, CH-1121, Genève 20, Switzerland/Suisse CH-1121

Phillips, S. D., K. R. Eberhardt, and B. Parry, "Guidelines for Expressing the Uncertainty Measurement Results Containing Uncorrected Bias," Journal of Research of the National Institute of Standards and Technology. Vol. 102, pg. 577. 1997.

Phillips, S. D., W. T. Estler, T. Doiron, K. R. Eberhardt, and M. S. Levenson, "A Careful Consideration of the Calibration Concept." Journal of Research of the National Institute of Standards and Technology. Vol. 106, pg. 1. 2001.

ISBN 0-7918-2772-0

9 780791 827727

L07902