A Unified Approach to Uncertainty for Quality Improvement

J E Muelaner[1], M Chappell[2], P S Keogh[1]
[1]*Department of Mechanical Engineering, University of Bath, UK*
[2]*MCS, Cam, Gloucester, UK*

Abstract

To improve quality, process outputs must be measured. A measurement with no statement of its uncertainty gives no meaningful information. The Guide to the expression of Uncertainty in Measurement (GUM) aims to be a universal framework for uncertainty. However, to date, industry lacks such a common approach. A calibration certificate may state an Uncertainty or a Maximum Permissible Error. A gauge study gives the repeatability and reproducibility. Machines have an accuracy. Processes control aims to remove special cause variation and to monitor common cause variation. There are different names for comparable metrics and different methods to evaluate them. This leads to confusion. Small companies do not necessarily have experts able to implement all methods. This paper considers why multiple methods are currently used. It then gives a common language and approach for the use of uncertainty in all areas of manufacturing quality.

1 Introduction

To improve quality process outputs must be measured. However, a measurement alone, without its uncertainty, is effectively meaningless since an error observed in a process may actually be an error of measurement. There is no way of knowing which is more likely, hence the basis of a quality system should be a consistent framework of uncertainty. This would allow statistical confidence to inform decisions.

The Guide to the expression of Uncertainty in Measurement (GUM) is intended to provide this framework [1]. It is established in National Measurement Institutes (NMI's) and calibration laboratories. It was intended to be applied universally in engineering, commerce and industry, specifically for "maintaining quality control and quality assurance in production". However, it has seen limited use in industry. Instead, a confusing array of methods and terms are used to quantify uncertainty.

Uncertainty is considered in different ways at different levels in the quality system. The highest levels, such as in NMI's and calibration laboratories, use the GUM. Instruments such as micrometers [2] are calibrated with stated *Uncertainties*. Less mature instruments may state *Maximum Permissible Error* (MPE). Examples include Coordinate Measurement Machines (CMM's) [3] and laser trackers [4]. Shop floor measurements are generally evaluated using a *Measurement Systems Analysis* (MSA) approach, as recommended in Six-Sigma [5]. This gives the *repeatability*, *reproducibility* and *accuracy* of measurements [6, 7]. When monitoring process outputs, yet another methodology, Statistical Process Control (SPC), is used. SPC uses control charts to identify *common cause* and *special cause variation,* and to check if a process is under *statistical control* [8].

A quality engineer must understand these approaches. Different terms may relate to equivalent quantities, with different methods used to evaluate them. This prevents the consistent use of uncertainty and optimal decisions may not be made. Small companies may not be able to afford experts in a number of similar disciplines. For them the situation is even more confusing. A large scale study was carried out with 235 small manufacturing companies. At least 2 days were spent working with each company to understand their measurement methods and requirements. None of these companies used an uncertainty based approach and few had an understanding of it. There was a general state of confusion as to the right process for product verification. The general feeling is summed up well by Mr Geoff Hayward in QEP's comments, "We would benefit from having one clearly defined process to follow using nice and simple engineering words to explain and not highly intellectual mathematic jargon. At the end of the day we are paid for the product we produce… there is no budget to employ maths geniuses to run simulations prior to manufacture". Hence, Industry needs a common uncertainty framework that applies to both measurements and processes.

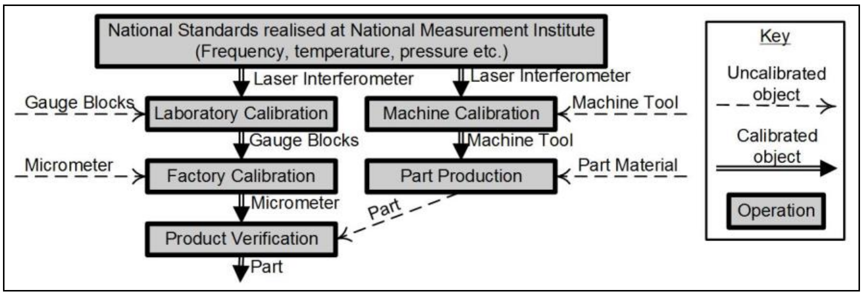

Figure 1: Simplified traceability chain part manufacture

Figure 1 shows a simplified traceability chain for a part produced to specified dimensions. In each operation a reference standard is used to calibrate an object.

This object transfers the dimensional information to the next operation. Types of operation include: calibration of instruments; calibration of machines; part production; and part verification. They are normally treated differently. However, the flow of dimensional information is broadly the same. Each operation is basically a calibration. It has a reference standard as the input and a calibrated object as the output. Each operation must: 1) Identify any influences on the output; 2) Make corrections for any measureable influences; and 3) Quantify the remaining uncertainty in order to state confidence. The part is not considered to be traceable until it has been verified because current machine tool calibrations are incomplete.

3 Current Approaches

The main current methods are within the GUM uncertainty framework, Measurement Systems Analysis (MSA) and Statistical Process Control (SPC). They share many common ideas but use different terms. The way they consider systemic effects has some differences. Perhaps the largest difference is whether they take a top-down or bottom-up approach. This relates to how they consider influence quantities. The following sections cover the different terms and methods.

3.1 Error, Accuracy and Uncertainty

An *error* of measurement is the result of a measurement minus a true value of the measurand. Error is sometimes confused with uncertainty, but it refers to a completely different concept. If a measurement is repeated then the true value will remain the same, but each result will be different. This is due to each measurement including different errors. We cannot know the true value and the error exactly. However, it is possible to estimate how close the true value is likely to be to the measurement result. This is the quantity expressed as accuracy or uncertainty.

In a general sense, accuracy and uncertainty refer to doubts about the validity of a measurement result. Accuracy is positive and uncertainty is negative. So if a measurement is more accurate its uncertainty is lower. The GUM defines *Uncertainty of Measurement* as the specific quantity which measures this. A *Standard Uncertainty* is this uncertainty given as a standard deviation. Uncertainty can be thought of as giving limiting values within which one can have confidence the true value lies. Typically, a single standard deviation does not give sufficient confidence. Assuming the uncertainty is normally distributed, limits at a higher confidence level can be calculated by multiplying the standard uncertainty by a *coverage factor*. This is equivalent to a z-score. Within MSA it is *Accuracy* that is the specific quantity. This can be confusing as in earlier standards accuracy meant only trueness. Accuracy is not normally used to calculate confidence limits.

The conventional view is that uncertainty arises from *error sources*. Within the GUM great care is taken not to confuse error with uncertainty. The term *influence quantities* is used for any 'quantity that is not the measurand but that affects the result of the measurement'. Within SPC these are called *factors*.

3.2 Random Effects

Random errors are those which change every time a measurement is repeated due to unpredictable effects, which may vary with time or position. GUM refers to the *random effects* which cause these errors. It notes that they cannot be compensated, but they may be reduced by averaging a number of measurements. MSA refers to *random error* while SPC uses *common causes* or, from Shewhart's original work, *chance causes*. All of these concepts are equivalent.

As discussed, individual errors are unknown, though random effects can be quantified by statistical analysis of repeated measurements. This is usually done by calculating the standard deviation. MSA refers to this as *precision*. The GUM calls it *random uncertainty*. Both GUM and MSA make the distinction between *repeatability* and *reproducibility*. Within SPC they are called *short-term variability* and *long-term variability*. Repeatability is the random uncertainty of results under the same conditions. Reproducibility is the random uncertainty under changed conditions. The conditions which may change include the operator, instrument, calibration and environment.

The GUM acknowledges the existence of random and systematic effects. But it recommends that uncertainties are not categorized in this way. This is because a random effect present when calibrating a reference will become a systematic effect when that standard is used to calibrate another instrument. To avoid this confusion the GUM identifies *type A uncertainties*. These are evaluated by statistical analysis of repeated observations. When evaluating a type A uncertainty there is one difference between a measurement and a production process. The same object may be measured multiple times, then the standard deviation calculated directly from the results. For a measurement the input is a physical object and the outputs are data. For a process this is reversed. The input of a process is data, the specification. A process then outputs a physical object. Therefore to determine the type A uncertainty for a production process each output must be measured.

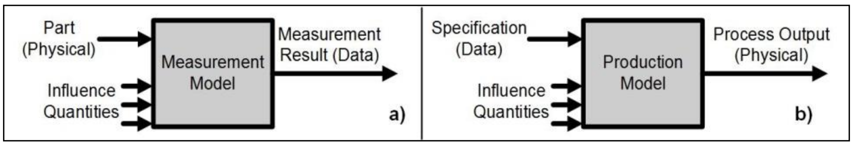

Figure 2: Comparison of Measurement and Production Process

3.3 Systematic Effects

Systematic errors are not random and may have a known cause. GUM refers to the *systematic effects* which cause these errors. They arise from the recognized effects of *influence quantities*. If the influence quantity can be quantified then a correction can be applied to compensate for the effect. In SPC these effects are now called *special cause variation*. Shewhart's original work refers to them as *assignable causes*. As discussed, the GUM doesn't categorise uncertainties by those arising from random or systematic effects. *Type B uncertainties* are those evaluated by means other than the statistical analysis of a series of observations. These often, although not necessarily, arise from systematic effects.

The GUM assumes that every effort is made to identify all significant systematic effects and then make corrections for them. SPC makes a similar assumption. SPC calls this a state of *statistical control* or a *stable process*. Only when a process is in control can statistical confidence limits can be validly applied. Therefore obtaining this state of control is central to SPC.

MSA takes a simple approach to systematic effects, by quantifying them and comparing the mean of many measurements with a reference value. The difference is the *bias*, also called *trueness*. MSA defines accuracy as the combination of precision and trueness. This assumes that the systematic effect of uncertainty in the reference standard is negligible. The GUM provides no such way of including uncorrected systematic effects in an uncertainty. Although the MSA approach does not depend on all systematic effects being compensated it is still recommended that any systematic effects are small compared to random effects [7].

3.4 Top-Down or Bottom-Up

A bottom-up approach starts with a model of the measurement. This is an equation giving the measurement result in terms of the influence quantities. The model is used to apply corrections and calculate uncertainty. A top-down approach starts by making a series of measurements. Statistical analysis is then used to identify any significant systematic effects and determine accuracy.

The GUM takes a bottom-up approach. It first requires a mathematical model of the measurement. The measurement process must include an estimate for each influence quantity. Estimates for significant systematic effects may be measurements or reference values. The model can then correct for these influences. For example, part temperature may influence a length measurement. In this case the part temperature would be measured and the resulting change in length calculated. Estimates for random effects are normally zero. The uncertainty of each influence quantity must also be calculated. Even if an influence quantity is estimated to be zero, it will have a non-zero uncertainty. The *Law of Propagation of Uncertainty* is then used to determine the combined uncertainty. It uses the mathematical model to combine the uncertainty in each influence quantity. This gives the uncertainty of the corrected measurement. Finding the uncertainty in this way requires an understanding of complex statistical assumptions. Expecting all industrial engineers to do this reliably is unrealistic. However, numerical methods such as Monte Carlo Simulation can also do this. With the right software this enables reliable calculation of uncertainty by a non-expert.

MSA takes a bottom-up approach. A standard part is measured, which has a reference value known with a small uncertainty. Repeated measurements are made. The difference between the mean of these measurements and the reference value is the bias. The standard deviation of them is the precision. If the bias is considered too large then efforts are made to identify and correct systematic effects. This approach is relatively simple. It may therefore appear that little can go wrong. However, its ease of use may encourage testing to be done without first considering all influence quantities. Therefore important influences may not be varied during the test. In this case systematic effects will not be seen in the test result. There may also be important influences which cannot be varied. This method is therefore prone to underestimate uncertainty. It can also be difficult to find a reference with a truly negligible uncertainty. In that case this influence may not be properly considered.

SPC follows a similar approach to MSA. It makes use of histograms and control charts to monitor random and systematic effects. When long term data are monitored in this way it is more reliable in detecting all influence quantities. This risks using non-conforming products to fix the process. It also depends on having long term data, which a short-run process will not have. It is better to identify effects from the outset.

There is also a risk with the GUM approach that the model has missed important influences. In industry there is often mistrust of this more theoretical approach. For this reason a comparison is recommended. A series of measurements should be made while varying as many influence quantities as possible. The model should then be used to estimate the uncertainty based on only those influences which were varied (GUM 3.4.2). This will result in a smaller combined uncertainty than that actually estimated for the measurement. The standard deviation in the measurement results should be of a similar value to the combined standard uncertainty calculated using the model.

1.6 Conformance and Capability

Knowing the uncertainty of a measurement or processes allows engineers to make better decisions. For example, does a measurement prove that a part is out of tolerance? Is a machine likely to produce parts within tolerance? These questions are answered very differently by the different methodologies. MSA and SPC consider whether a measurement or production processes is *capable*. In this sense capable means that the accuracy or variation is small compared to the tolerance of the part. If the process is found to be capable then the uncertainty is not considered any further.

For example, a tolerance states that a dimension must be between 8.9 mm and 9.1 mm. A measurement result of 8.9 mm is regarded as within specification. An uncertainty based approach is given in ISO 14253 [9]. This states that to prove conformance a tolerance by the uncertainty of measurement must be reduced. The confidence level of the uncertainty is used is the confidence that exits for that part to be conforming. In this example the measurement result would not prove conformance. It would show an equal chance of conformance or non-conformance. It is assumed that the uncertainty of the measurement is 0.01 mm at 95% confidence. The smallest measurement result which would prove conformance with 95% confidence would therefore be 8.91 mm. In fact this gives rise to a basic statistical error. The confidence level for an expanded uncertainty gives the region either side of the measurement result within which the true value is likely to lie. In statistical terms this would be the probability for a two-tailed test. But we are concerned with a measurement result which is close to a tolerance limit. If the true value is closer to the nominal value then this is not a problem. However, if the true value is in the direction of the limit then it may actually exceed it. So in statistical terms we need the probability for a one-tailed test! If ISO 14253 is applied directly it will therefore be overly cautious. In this case the extent to which a tolerance is reduced by the uncertainty gives an idea of capability.

To fully apply uncertainty in proving part conformance and process capability they must be considered together. When a process produces a part and a measurement verifies it, there are four possible outcomes. A) The part is in tolerance and the measurement correctly proves this. B) The part is out of tolerance and the measurement correctly proves this. C) The part is in tolerance but an error of measurement means that it is rejected. D) The part is out of tolerance but an error of measurement means that it is passed. Case A is the desired outcome. B, C and D are each a different type of quality issue. Each will have a different cost. The occurrence rate of each may be obtained by combining the uncertainty in the process with the uncertainty in the measurement. This has no analytical solution, but can be calculated numerically. Three variables can then be optimized for cost [10]. These are the uncertainty in the process, the uncertainty in the measurement and the limits set for part acceptance. This is the ultimate aim of quality management.

4 Unified Approach

This paper suggests that industry moves to a unified uncertainty based approach. This would use the same methods to determine the uncertainty of production machines, processes and measurements. An initial foundation for such an approach is a standardised vocabulary. Table 1 gives the most important terms in this vocabulary. The remainder of this paper uses this preferred vocabulary. This predominantly represents a restricted subset of the VIM terminology with the addition of the concepts of statistical control, conformance limits, etc.

Table 1: Unified Vocabulary

<table>
<tr><th>Preferred Terms</th><th colspan="2">Usage and related non-preferred terms</th></tr>
<tr><td>Process</td><td colspan="2">General term for measurements, measurement processes, machines and production processes</td></tr>
<tr><td>Uncertainty</td><td>General concept of doubt about the output of a process</td><td rowspan="4">Related: Accuracy</td></tr>
<tr><td>Standard Uncertainty</td><td>Quantity representing uncertainty as a standard deviation</td></tr>
<tr><td>Combined Standard Uncertainty</td><td>Standard uncertainty in process output, determined by combining standard uncertainty in each influence quantity</td></tr>
<tr><td>Expanded Uncertainty</td><td>Standard uncertainty multiplied by a coverage factor for a confidence level</td></tr>
<tr><td>Confidence Level</td><td colspan="2"></td></tr>
<tr><td>Confidence Limits</td><td colspan="2"></td></tr>
<tr><td>Coverage Factor (k)</td><td colspan="2">z-score (two-tailed)</td></tr>
</table>

Single-Sided Coverage Factor (k_1)	z-score (one-tailed)
Influence Quantity	Related: Input Quantity; Error Source; Factor
Influence Model	New term suggested for the mathematical model. It gives the process result in terms of the influence quantities and is used to determine uncertainty.
Error	The unknowable difference between a result and the true value
Random Effects	Related: Random Error; Common Causes; Chance Causes
Random Uncertainty	Related: Precision; Common Cause Variation; Chance Cause Variation
Repeatability	Related: Short-term variability
Reproducibility	Related: Long-term variability
Type A Uncertainty	
Systematic Effects	Related: Special Causes; Special Cause Variation; Assignable Causes; Systematic Uncertainty
Bias	Related: Trueness
In Statistical Control	A corrected result or process with negligible uncorrected systematic effects. Related: Corrected result; Stable process
Type B Uncertainty	
Tolerance	
Specification limits	
True reject rate	The percentage of non-conforming parts correctly failed by verification
False reject rate	The percentage of conforming parts incorrectly failed by verification
False pass rate	The percentage of non-conforming parts incorrectly passed by verification
Cost of reject	The cost of rejecting a part because it is believed to be non-conforming
Cost of false pass	The cost of allowing a non-conforming part to reach the customer
Conformance limits	Tightened specification limits to prevent

For Industry to move to an uncertainty based approach it will require more than common language. Another barrier is that the GUM assumes a fully corrected result. Industry measurements must have an appropriate level of uncertainty with minimum cost. Therefore some known significant systematic effects may not be corrected if it is not economical to do so. This can be easily overcome by a slightly more relaxed application of the GUM. Uncertainties resulting from systematic effects are combined with random uncertainties. For example, the uncertainty in a reference standard used for calibration. So the same approach can be taken to uncorrected systematic effects without further complication.

The bottom-up method used to determine uncertainty is a more serious barrier. It is challenging to expect engineers to create a mathematical model for each process. For them to then analytically combine the uncertainties is even more difficult. For specialist metrologists there is a significant risk of introducing human error. For general industrial engineers the method could be totally impractical. However, for many measurements standard models may be used. These should be made available by manufactures of instruments and machines, even if just in black-box form.

Analytical methods are not necessary. Monte Carlo Simulation (MCS) can replace them. It is a robust way to find the uncertainty for any model with no user input. It is also intuitive enabling non-technical users to better understand uncertainty.

Algorithms are required to provide these standard models and apply MCS to them. Software could give information on underlying principles while automatically simulating the combination of uncertainty. This would promote understanding while giving robust mathematical solutions. The software could prompt users to enter worst case estimates for influence quantities not measured in the process. Sensitivity analysis would show which influences are significant. The user could then obtain improved estimates for these.

Checks must validate both the process and the model used for uncertainty estimation. As discussed above, this involves reproducibility studies which vary as many influence quantities as possible. The results are compared with the model prediction for these reproducibility conditions. Such tests should be carried out regularly on an ongoing basis. When tests identify uncorrected influence quantities these should be added to the influence model. Some of the graphical methods of SPC could be useful in this stage.

The paper has thus far considered a mainly linear traceability chain. Figure 1 shows this simplified view. In reality the situation is somewhat more complex. Figure 3 shows a part of the chain in more detail. The machine tool is first calibrated using a laser to make corrections for known influence quantities. Other instruments may also be used to correct other influences. Parts are then produced and measured. It is their variability which is used to validate the uncertainty estimation. It may also be possible to use information about the parts to provide ongoing corrections for influences which drift over time. For example, a drift in scaling could be easily detected and corrected using measurements of parts produced. More complex model based corrections may be possible by fitting observed errors to the model to determine which influence quantities would give these errors in the output. Such an approach may enable self-calibrating machines.

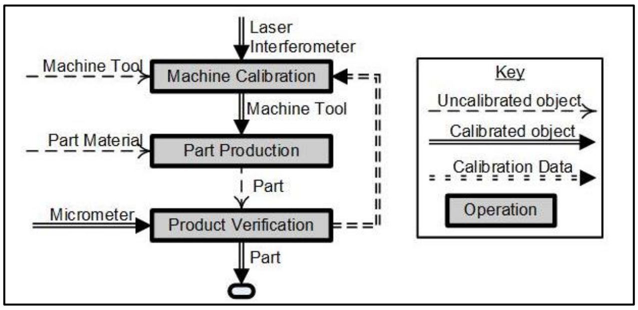

Figure 3: Production Process with Feedback from Part Verification

The complexity of real industrial processes makes manual calculation increasingly difficult. Examples include thermal expansion and coordinate measurements. With a length measurement thermal expansion may be considered as a simple scaling for the average change in temperature. With more complex parts it may be necessary to consider temperature variation over the part. Thermal expansion then results in bending and twisting. This requires much more complex models for compensation and uncertainty estimation. Similarly when the coordinates of a feature on a part is measured with respect to multiple datums this results in a complex influence model. These considerations further highlight the need for software to support uncertainty calculations.

7 Conclusions and Further Work

Despite differences of language current approaches have a great deal in common. The first step in moving to a unified uncertainty based approach is to standardise this language. A suggested vocabulary is provided. Other minor obstacles can be easily overcome through an improved understanding and slight relaxation of the GUM approach. These include carrying out before-the-fact uncertainty evaluation and including uncorrected systematic uncertainty.

The main barrier to the adoption of uncertainty in industry is the bottom-up approach. This involves two difficulties. The first is the requirement to obtain mathematical models for processes. This could be greatly assisted if manufacturers provided models for instruments and machines. Standard equations for simple processes may also be made available. The second difficulty is in the analytical methods of combining uncertainty. This can be simply overcome by using Monte Carlo Simulation. These solutions are available now for some applications and with knowledge of computational methods. General purpose algorithms could make it fully applicable.

Algorithms should also support model checks against reproducibility studies, thermal deformation and complex datum structures. Furthermore they must enable the optimization of process uncertainty, measurement uncertainty and conformance limits. With this final step the unified approach to uncertainty will give significant improvements in cost and quality.

Acknowledgements

The research described herein has been carried out as part of the EPSRC, Light Controlled Factory Project (grant No. EP/K018124/1) at the Laboratory for Integrated Metrology Applications (LIMA) in the Mechanical Engineering Department of the University of Bath.

References

1. BSI, *General Metrology - Part 3: Guide to the expression of uncertainty in measurement (GUM)*, in *PD 6461-3*. 1995.
2. ISO. *Geometrical product specifications (GPS) - Dimensional measuring equipment: Micrometers for external measurements - Design and metrological characteristics*. 3611 2010 [cited.
3. BSI – British Standards Institution, *Geometrical Product Specifications (GPS) - Acceptance and reverification tests for coordinate measuring machines (CMM) - Part 2: CMMs used for measuring size*, in *BS EN ISO 10360-2:2002*. 2002.
4. ISO, *Geometrical product specifications (GPS) - Acceptance and reverification tests for coordinate measuring systems (CMS) - Laser trackers for measuring point-to-point distances*, in *10360-10*. 2016.
5. ISO, *ISO 13053-1: Quantitative methods in process improvement — Six Sigma Part 1: DMAIC methodology*. 2011.
6. ISO. *5725: Accuracy (trueness and precision) of measurement methods and results*. Part 1: General principles and definitions 1994 [cited.
7. Auromotive Industry Action Group, *Measurement Systems Analysis*. 4th edition ed. 2010.

8. ISO, *Guidelines for implementation of statistical process control (SPC) - Part 1: Elements of SPC*, in *11462-1*. 2001.
9. BSI, *Geometrical Product Specifications (GPS) — Inspection by measurement of workpieces and measuring equipment — Part 1: Decision rules for proving conformance or non-conformance with specifications*, in *BS EN ISO 14253-1:1999*. 1999.
10. Forbes, A.B., *Measurement uncertainty and optimized conformance assessment.* Measurement, 2006. **39**: p. 808–814