THE RAMIFICATIONS FOR DESIGN PERFORMANCE OF MAX METAL MACHINING PRACTICES IN HIGH VALUE, HIGH PRECISION APPLICATIONS

Mr. G. May, CEng, FIMechE
(Rolls-Royce plc, United Kingdom);

Dr. P. Rowe,
(Bourton Group Ltd, United Kingdom);

Dr. A. Kumar Sinha, CEng, FRAeS,
(Rolls-Royce plc, United Kingdom)

1. Introduction

In attempting to better manage and reduce the effective unit cost of component manufacture there is both a widespread and a natural tendency to establish processes that promote a predictable degree of re work and avoidance of scrap. This is particularly the case for high value, high precision components, that involve material removal as part of the overall manufacturing process and where the cost of scrapping a component as defective (if too much material is removed) is significant in terms of the amount of wasted time, effort and raw materials.

In the Gas Turbine industry this is particularly true where the use of expensive, hard-to-source high temperature alloys is common and individual components can therefore incur cost in the thousands, if not tens of thousands of pounds if scrapped. The manufacturing practices adopted in such cases can be descriptively termed *Max. Metal Machining* (MMM), in that they typically involve an incremental approach to producing components that are within specification limits: after an initial amount of material is removed, the component's dimensions are measured. If they fall inside the specification limit, the component is accepted as conforming to the specification, otherwise an additional material removal pass and subsequent re measurement is performed, which process is repeated until the component can be accepted.

While the practice of MMM appears a logical one in relation to the narrow focus on component unit cost, there is also the potential for a significant negative impact on a product's ability to satisfy its function and other requirements (such as weight and life targets), since the results of MMM will be a highly skewed distribution of component dimensions towards one end of the specification limits, rather than a conformance to the design intent of most dimensions being close to the target nominal and very rarely coming close to either specification limits. Design predictions for nominal performance – and the robustness of such performance – that may have been based on the expectation that the manufacturing process behaves as such will therefore be inaccurate.

A further impact of adopting MMM is apparent when the presence of *Measurement Systems Variation* (MSV) is recognised and accounted for: collected data will be significantly distorted due to the (repeated) measurement process, and due to the localisation of measurements about one end of the specification limits there will likely be a significant rate of miss classification of measured components – specifically those that are in fact measured and classified as being conforming, when in fact they are actually outside the specification limits and should have either been rejected or subject to further re work.

This paper, using the example of a generic bearing housing assembled with an interference fit to a shaft, highlights the impact of adopting MMM on both the typical behaviour and robustness of the assembly through the use of *Monte Carlo Simulation* (MCS) to predict the resulting mean, variation and distribution in interference fit, based on distributional models of the bearing housing and shaft diameters, derived from synthesised measurements. Additionally, the susceptibility of this manufacturing approach to the effects of MSV for such an assembly, again using MCS, will also be demonstrated. For comparison, the same evaluations are completed for the same assembly when produced from processes that are conducted according to an *On Target with Minimal Variation* (OTWMV) approach in which the machining processes for both bearing housing and shaft are centred on the (design intent) nominal values and assumed to be in *statistical control* (the processes are stable and predictable, each with constant mean and variance).

2. Process Variation and Measurement Variation

It is certainly true that no real-world process can be said to be perfect, since no process, when repeated, will produce exactly the same outcome, time after time. In the context of Manufacture this is recognised in the form of specifying both a desired nominal value and upper and lower tolerance limits (specification limits), which therefore allow us to classify the outcome of a process as being conformant to the specification (within the defined limits) or else non-conformant (outside the defined limits). This binary classification essentially means that the specification limits are in general treated as "goal posts", i.e. all parts produced that are conformant are assumed to be "equally good" and all parts produced that are non-conformant are treated as "equally bad". As we will go on to discuss later in this paper, from the viewpoint of delivering consistent product performance, this is assumption does not in fact prove to be true.

The means of classification of parts necessarily involves an act of measurement – and since this is a process in its own right, it is inevitable that measurements too are variable: we cannot expect to report exactly the same measurement if either a single person (or operator) measures the same part multiple times (the measurement is not perfectly *repeatable*), or else multiple persons (operators) measure the same part once (the measurement is not perfectly *reproducible*).

A *Gauge R&R study* is specifically designed for the purpose of quantifying the variation present in a measurement process, and also how much of the random error can be categorised as *Repeatability* and *Reproducibility*. Such studies also produce two overall quality metrics that can be used to assess whether or not the particular measurement system is "fit for purpose"

- In the context of understanding how well a measurement system can correctly differentiate one part from another the *Percentage Study* (*%Study*) metric is relevant. This metric is essentially the ratio of the "width" of the measurement error distribution (which is assumed normal) and the "width" of the observed variation in the measurements taken from all parts measured during the study, expressed as a percentage:

  $$\%Study = \frac{6\sigma_{GRR}}{6\sigma_{tot}} \times 100$$

  where σ_{GRR} is the standard deviation of the measurement system and, σ_{tot} is the standard deviation of all observed measurements taken (i.e. this includes the true part-to-part variation and the measurement system variation too).

- In the context of classifying parts as being either conformant or non-conformant to the specification limits, the *Percentage Tolerance* (*%Tol*) metric is relevant. This metric is essentially the ratio of the "width" of the measurement error distribution (which again is assumed normal) and the specification width, expressed as a percentage:

    - For 1-sided specifications:

    $$\%Tol = \frac{1}{2}\frac{6\sigma_{GRR}}{|\bar{y} - SL|} \times 100$$

    where σ_{GRR} is the standard deviation of the measurement system, $\bar{y}$ is the mean of the measured data and *SL* is the single specification limit.

    - For 2-sided specifications:

    $$\%Tol = \frac{6\sigma_{GRR}}{|USL - LSL|} \times 100$$

    where again σ_{GRR} is the standard deviation of the measurement system, and *USL* and *LSL* are the upper and lower specification limits respectively.

Notwithstanding that an acceptable value depends on the situation and criticality of the measurement, as a rule of thumb values for both *%Study* and *%Tol* of 10% or less indicate an excellent measurement system, whereas values greater than 30% indicate that the measurement system is not fit for purpose. Values of the order of 20-25% are not untypical.

3. Misclassification of Measured Parts

Given the presence of variability in measurements there are four possible outcomes when a measurement is taken in terms of part classification, namely:

- Accept a good part (the true value of the part is conformant to the specification, and the measured value, while likely different, is also conformant).

- Reject a bad part (the true value of the part is non conformant to the specification, and the measured value, while likely different, is also non conformant).

- Reject a good part (the true value of the part is conformant to the specification, but the measured value is non conformant).

- Accept a bad part (the true value of the part is non conformant to the specification, but the measured value conformant).

The first two of these outcomes both correctly classify parts as conformant and non-conformant respectively, whereas the last two outcomes are cases where the parts have been misclassified. This idea is illustrated in Figure 1:

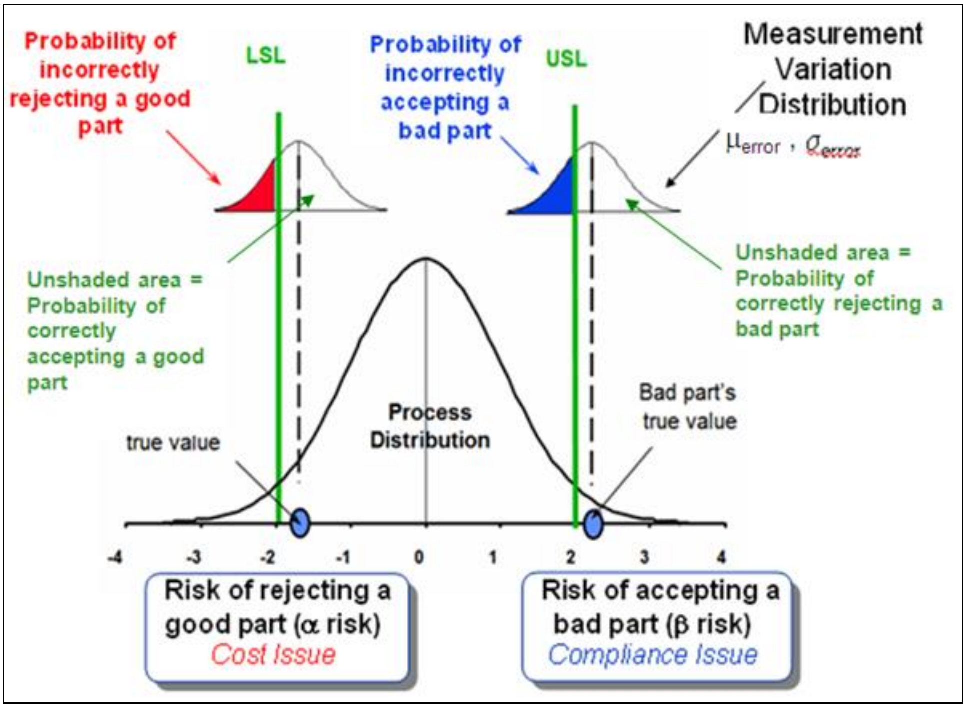

Figure 1 Classification and Misclassification of Parts due to Measurement System Variation

From Figure 1 it is evident that the problem of misclassification is exacerbated if the probability of parts being near to the specification limits is high (because the process capability, usually expressed using a capability index such as C_{pk}, is low), and/or if the measurement error is large relative to the specification width. C_{pk} is calculated as follows:

$$C_{pk} = \frac{k}{3\sigma_{ST}}$$

Where k is the smaller of the two results, $USL - \mu$ and $\mu - LSL$; μ is the process mean; σ_{ST} is the short-term (inherently common cause) process standard deviation and *LSL* and *USL* are the lower and upper specification limits, respectively.

Process capability metrics such as C_{pk} express the expected defect rate (under certain distributional assumptions), for example: for a process that produces a normally distributed characteristic of interest, a *Cpk* value of 1.0 implies a defect rate of approximately 2700ppm, (for a bilateral specification) whereas a C_{pk} value of 2.0 implies a defect rate of approximately 1ppb (again for a bilateral specification). The latter case is commonly referred to as a *Six Sigma* level of capability. Processes with C_{pk} values less than 1.0 are at significant risk of part misclassification.

The rates of each type of misclassification can be calculated – and the means of such calculation is provided in software packages such as Minitab[1] when using it to define and analyse the results of Gauge R&R studies. The calculation assumes a bivariate normal distribution for the underlying and the observed data (i.e. a normal error distribution). The same results can be achieved using *Monte Carlo Simulation* (MCS), but this does not constrain the distributional form of the underlying data or measurement error distribution. In this paper we will review both methods.

4. Calculating Misclassification Rates from a Gauge R&R Study

If we define the region within the specification limits as *R*, the true part's dimension is a random variable *X* and its measured value is a random variable *Y* then due to the nature of the problem *X* and *Y* are not independent. We can, therefore, quantify misclassification rates via both *joint probabilities* and *conditional probabilities*:

- **Joint probabilities** will tell us what proportion of the total population will fall into the two categories: parts that are accepted as good when they are in fact bad and parts that are rejected as bad when they are in fact good.
- **Conditional probabilities** will further tell us the proportion of good parts that will be incorrectly rejected as bad and the proportion of bad parts that will be incorrectly accepted as good.

<u>The Joint Probability that a Part is Bad but is Accepted as Good</u>

For a bivariate distribution it is the case that the probability of being in a region bounded by a range of *X* and a range of *Y* is:

(1)
$$P(a \le X \le b, c \le Y \le d) = F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$$

To understand this consider a rectangular region bounded by (*a*,*b*) on the *x*-axis and (*c*,*d*) on the *y*-axis. For the case under consideration it is not difficult to see that the joint probability that a part is bad but is misclassified and so is accepted as good is:

(2)
$$P(X \notin R, Y \in R) = P(X < L, L \le Y \le U) + P(X > U, L \le Y \le U)$$

So, referring to Equation (1), the first part of the right hand side of Equation 2 has: *a* = -∞, *b* = *L*, *c* = *L* and *d* = *U* and the second part has: *a* = *U*, *b* = ∞, *c* = *L* and *d* = *U*.

Now, since a value of *X* less than minus infinity is impossible, regardless of the value of *Y*, and vice versa, then:

$$\lim_{x \to -\infty} F_{X,Y}(x,y) = \lim_{y \to -\infty} F_{X,Y}(x,y) = \lim_{x \to -\infty, y \to -\infty} F_{X,Y}(x,y) = 0$$

Also, since $F_{X,Y}(\infty,y) = F_Y(y)$ and $F_{X,Y}(x,\infty) = F_X(x)$ (viz. the marginal cumulative distribution functions of *X* and *Y*), so Equation (2) becomes:

$$P(X \notin R, Y \in R) = \{F_{X,Y}(L,U) - F_{X,Y}(-\infty,U) - F_{X,Y}(L,L) + F_{X,Y}(-\infty,L)\} + \{F_{X,Y}(\infty,U) - F_{X,Y}(U,U) - F_{X,Y}(\infty,L) + F_{X,Y}(U,L)\}$$

$$P(X \notin R, Y \in R) = \{F_{X,Y}(L,U) - F_{X,Y}(U,U) - F_{X,Y}(L,L) + 0\} + \{F_Y(U) - 0 - F_Y(L) + F_{X,Y}(U,L)\}$$

(3)
$$P(X \notin R, Y \in R) = F_Y(U) - F_Y(L) - F_{X,Y}(U,U) + F_{X,Y}(L,U) + F_{X,Y}(U,L) - F_{X,Y}(L,L)$$

<u>The Joint Probability that a Part is Good but is Rejected as Bad</u>

Similarly to Equation (2), the joint probability that a part is good but is misclassified and is rejected as bad is:

(4)
$$P(X \in R, Y \notin R) = P(L \le X \le U, -\infty < Y < L) + P(L \le X \le U, U < Y < \infty)$$

Again using the same techniques as above:

$$P(X \in R, Y \notin R) = \{F_{X,Y}(U,L) - F_{X,Y}(L,L) - F_{X,Y}(U,-\infty) + F_{X,Y}(L,-\infty)\} + \{F_{X,Y}(U,\infty) - F_{X,Y}(L,\infty) - F_{X,Y}(U,U) + F_{X,Y}(L,U)\}$$

$$P(X \in R, Y \notin R) = \{F_{X,Y}(U,L) - F_{X,Y}(L,L) - 0 + 0\} + \{F_X(U) - F_X(L) - F_{X,Y}(U,U) + F_{X,Y}(L,U)\}$$

(5)
$$P(X \in R, Y \notin R) = F_X(U) - F_X(L) - F_{X,Y}(U,U) + F_{X,Y}(L,U) + F_{X,Y}(U,L) - F_{X,Y}(L,L)$$

In Minitab the joint probability density distribution is assumed to be a *bivariate normal distribution*, with marginal distributions also therefore being normal:

(6)
$$X \sim N(\mu, \sigma_{pp}^2) \quad \text{and} \quad Y \sim N(\mu, \sigma_{tot}^2)$$

Where σ^2_{pp} is the variance of the true values of the dimension of the parts concerned and $\sigma^2_{tot} = \sigma^2_{pp} + \sigma^2_{GRR}$ (in our terminology $\sigma_{tot} = \sigma_Y$ and $\sigma_{pp} = \sigma_X$).

Now, the general multivariate normal joint probability density function $f_{X,Y}(x,y)$ is given by the following equation:

(7)
$$f_{X,Y}(x,y) = \frac{1}{\sqrt{(2\pi)^d \det \underline{\Sigma}}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x} - \underline{\mu})\right]$$

Where the vector $\underline{x}$ is the matrix of random variables and the vector $\underline{\mu}$ is the matrix of mean values for each random variable.

We are interested in the bivariate case, for which $d = 2$. In this case $\underline{x} = (X\ Y)^T$, with X and Y defined as in Equation (6). Also, the parts to be measured have a population mean μ and we are not considering bias in the measurement system at this juncture, so $\underline{\mu} = (\mu\ \mu)^T$.

In the bivariate normal case:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\left(\frac{g(x)-h(x,y)+l(y)}{\sigma_x^2\sigma_y^2\left(1-\rho^2\right)}\right)\right]$$

Where $g(x) = \sigma_y^2\left(x-\mu_x\right)^2 = \sigma_{tot}^2\left(x-\mu\ \right)^2$

And $h(x,y) = 2\rho\sigma_x\sigma_y\left(x-\mu_x\right)\left(y-\mu_y\right) = 2\sigma_{pp}^2\left(x-\mu\ \right)\left(y-\mu\ \right)$

And $l(y) = \sigma_x^2\left(x-\mu_y\right)^2 = \sigma_{pp}^2\left(x-\mu\ \right)^2$

The corresponding joint cumulative probability distribution is given by:

(8) $$F_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{\det\underline{\Sigma}}}\left(\int_{u_1=-\infty}^{y}\int_{u_2=-\infty}^{x}\exp\left(-\frac{1}{2}\underline{w}^T\underline{\Sigma}^{-1}\underline{w}\right)\right)du_1du_2$$

The vector $\underline{w} = (\mu_1 - \mu,\ \mu_2 - \mu)^T$ and $\underline{\Sigma}$ is the variance covariance matrix, which is a symmetric, positive-definite matrix that has dimension $d\times d$.

For the bivariate case:

(9) $$\underline{\Sigma} = \begin{pmatrix} V(X) & Cov(X,Y) \\ Cov(X,Y) & V(Y) \end{pmatrix}$$

In general, $Cov(X,Y) = \rho\cdot\sigma_x\cdot\sigma_y = \rho\cdot\sigma_{pp}\cdot\sigma_{tot}$, where ρ is Pearson's linear correlation coefficient. For the particular case we are interested in, $V(X) = \sigma^2_{pp}$ and $V(Y) = \sigma^2_{tot}$. Also, for a bivariate normal distribution it is the case that:

$$\sigma_{Y|X}^2 = \sigma_Y^2\left(1-\rho^2\right) \Leftrightarrow \sigma_{tot|pp}^2 = \sigma_{tot}^2\left(1-\rho^2\right)$$

$$\Rightarrow \frac{\sigma_{tot|pp}^2}{\sigma_{tot}^2} = \left(1-\rho^2\right) \Leftrightarrow \rho^2 = \frac{\sigma_{tot}^2 - \sigma_{tot|pp}^2}{\sigma_{tot}^2}$$

$$\Rightarrow \rho^2 = \frac{\sigma_{tot}^2 - \sigma_{GRR}^2}{\sigma_{tot}^2}$$

The final expression above follows from the fact that, given a part's true value, its measured value is determined by the variation due to repeatability and reproducibility errors. Also, using the relationship $\sigma^2_{tot} = \sigma^2_{pp} + \sigma^2_{GRR}$, and, since standard deviations must always be positive, we can simplify the above expression to:

(10) $$\rho^2 = \frac{\sigma^2_{pp}}{\sigma^2_{tot}} \Leftrightarrow \rho = \frac{\sigma_{pp}}{\sigma_{tot}}$$

The expression for ρ in Equation (10) also makes physical sense:

- If there is *no* random measurement error from repeatability and reproducibility sources then *%Study* = 0 and $\sigma_{tot} = \sigma_{pp}$, so $\rho = 1$; i.e. the measured part dimension will be the same as its actual dimension (assuming no bias).
- If there is *only* random measurement error from repeatability and reproducibility sources then *%Study* = 100%, $\sigma_{GRR} = \sigma_{tot}$ and $\sigma_{pp} = 0$, so $\rho = 0$; i.e. the measured dimension bears no (linear) relation to the actual dimension.

So, since $Cov(X,Y) = (\sigma_{pp}/\sigma_{tot})\bullet\sigma_{pp}\bullet\sigma_{tot} = \sigma^2_{pp}$ then:

(11) $$\underline{\Sigma} = \begin{pmatrix} \sigma^2_{pp} & \sigma^2_{pp} \\ \sigma^2_{pp} & \sigma^2_{tot} \end{pmatrix}$$

It follows that the determinant of $\underline{\Sigma}$ is: det $\underline{\Sigma} = \sigma^2_{pp}\cdot\sigma^2_{tot}\cdot(1 - \rho^2)$.

The *marginal* cumulative distribution functions of *X* and *Y* are simply those for univariate normal distributions, e.g. for *X*:

(12) $$F_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}}\left(\int_{u=-\infty}^{x} \exp\left[-\frac{(u-\mu)^2}{2\sigma_x^2}\right]du\right)$$

Where, again, *X* is the random variable representing the true part-to-part variation, and so $\sigma_X = \sigma_{pp}$. The expression for $F_Y(y)$ follows directly, from symmetry.

The Conditional Probability that a Part is Bad but is Accepted as Good

The probability that a part is measured as good to specification by the measurement process, when in fact it is given that it is defective is given directly by the law of conditional probability as:

(13) $$P(Y \in R \mid X \notin R) = \frac{P(X \notin R, Y \in R)}{P(X \notin R)}$$

Where $P(X \notin R) = 1 - P(X \in R)$ and $P(X \in R) = F_X(U) - F_X(L)$.

The Conditional Probability that a Part is Good but is Rejected as Bad

The probability that a part is measured as defective by the measurement process, when in fact it is good is given is given similarly to Equation (13) as:

(14) $$P(Y \notin R \mid X \in R) = \frac{P(X \in R, Y \notin R)}{P(X \in R)}$$

Example Calculation of Misclassification Rates from Joint and Conditional Probabilities

In order to demonstrate the use of the joint and conditional probabilities arrived at in Equation (3), Equation (5), Equation (13)and Equation (14), we will use the simple example of a generic component that has a key dimension with a specification range of 2.0in to 2.6in and a desired nominal of 2.3in.

Firstly, a (normal) Capability Analysis on a production run conducted using Minitab revealed that the process has a C_{pk} value of 0.5 (very poor) with an expected defect rate of 12.44%. Inherent in the measurement data used in this analysis is random *repeatability and reproducibility* error, and hence there will be an associated rate of misclassification of parts during production.

A Gauge R&R study was next defined in Minitab with a representative sample of ten parts selected randomly from the production population and three operators selected at random. Each part was measured twice each by every operator in a randomised sequence. The results of Minitab's analysis were as follows:

- *%Study* = 32.66%
- *%Tol* = 66.61%
- S_{pp}, an *estimate* of σ_{pp} from the sample = 0.19278in
- S_{tot}, an *estimate* of σ_{tot} from the sample = 0.20397in
- S_{GRR}, an *estimate* of σ_{GRR} from the sample = 0.06661in

In other words, almost one third of the observed variation between parts was actually due to measurement variation instead (the *%Study* value), and almost two thirds of the specification width for this dimension is taken up by the measurement system variation (the *%Tol* value). So, in this case the measurement system would be classified as not fit for purpose, since both the *%Study* and the *%Tol* figures are above 30%.

Minitab also reported the probabilities of misclassification as:

- Joint Probabilities that:
  - A part is bad and accepted = 2.4%
  - A part is good and is rejected = 4.5%
- Conditional Probabilities of:
  - False Acceptance = 19.7%
  - False Rejection = 5.1%

In other words: 2.4% of production is wrongly accepted by the measurement system; 4.5% of production is wrongly rejected by the measurement system; of all defective parts, 19.7% will be wrongly accepted and of all good parts, 5.1% of them will be wrongly rejected. The first two results are given by Equation 3 and Equation 5 and the last two are given by Equation 13 and Equation 14.

Calculation and Interpretation of Misclassification Rates from Monte Carlo Simulation

For the purposes of comparison normal distributions for *X* (the true dimension) and *GRR*, the random R&R error, were defined as: $X \sim N(2.3075, 0.19278^2)$ and $GRR \sim N(0, 0.06661^2)$.

Values for both *X* and *GRR* were sampled by the MCS algorithm as implemented in Crystal Ball[2], a plug in extension to Excel[3]. Each pair of values was summed to give *Y*, where $Y = X + GRR$, and the result classified as belonging to one of four categories, based on the same specification limits of 2.0in to 2.6in as stated previously:

1. *X* was inside the specification but *Y* was outside the specification ("failed good").
2. *X* was outside specification but *Y* was inside specification ("passed bad").
3. *X* was outside specification, as was *Y* ("failed bad").
4. *X* was inside specification, as was *Y* ("passed good").

The categorisations for a sample of 10^6 such trials are as shown in Figure 2:

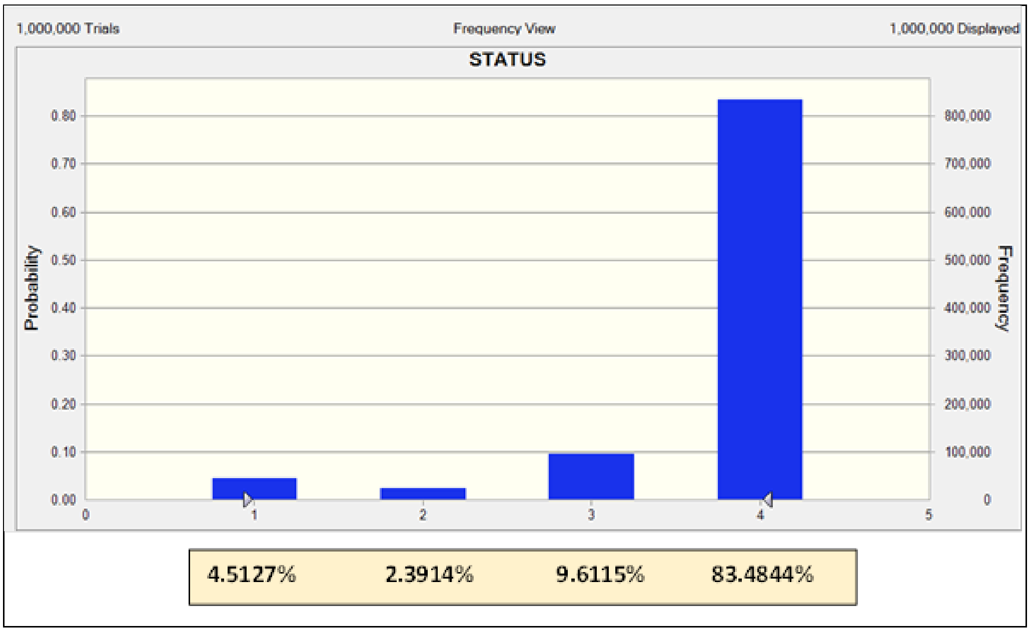

Figure 2 Classification and Misclassification of Parts from Monte Carlo Simulation

The equivalent of the conditional probabilities as calculated by Minitab can also be derived from these results, as follows:

- The False Accept (proportion of Bad parts falsely accepted as Good) value is:

$$\frac{\text{"passed bad"}}{\text{"passed bad"}+\text{"failed bad"}} = \frac{2.3914}{2.3914+9.6115} = 19.924\%$$

- The False Reject (proportion of Good parts falsely rejected as Bad) value is:

$$\frac{\text{"failed good"}}{\text{"failed good"}+\text{"passed good"}} = \frac{4.5127}{4.5127+83.4844} = 5.128\%$$

As can be seen, the values for categories 1 and 2 agree closely with the joint probabilities given by Minitab, as do the derived values for the conditional probabilities (allowing for rounding of figures in Minitab and sampling variation in MCS).

<u>Conditional Probabilities for Specific Dimensional Values for Varying Measurement Capabilities</u>

For a given value of a part dimension, the probability of misclassification depends upon its position relative to the specification limits and the %*Tol* figure. We shall now consider the case of machining an internal bore diameter on a bearing housing, and so calculate the effect of different %*Tol* values on the probability of misclassification for a range of bore diameter values both inside and outside the design specification of: lower limit = 1.2497in and upper limit = 1.2503in. The desired nominal diameter = 1.2500in.

As before, we can work out the probability of both types of misclassification (rejecting a good part as bad and accepting a bad part as good) using conditional probabilities as defined by Equation (13)and Equation (14). They can be calculated easily using the Excel formula NORMSDIST(*Z*), where *Z* is the standard normal deviate that gives the number of standard deviations σ_{GRR} corresponding to the difference between the actual diameter and the appropriate specification limit. The results for %*Tol* values of 10%, 30% and 100% (representing good, poor and extremely poor measurement systems) are shown in Figure 3 and Figure 4.

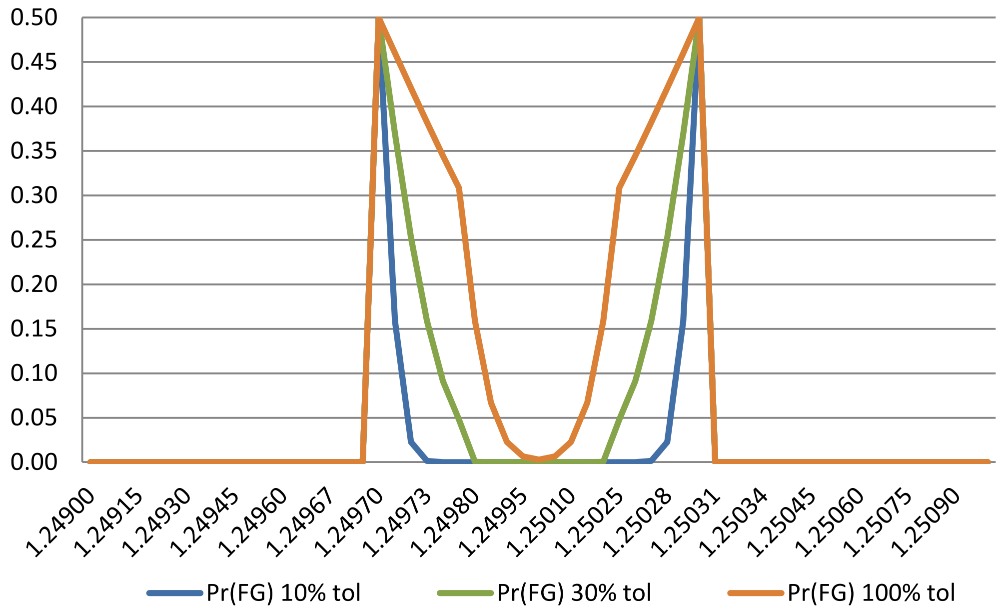

Figure 3 Probability of Failing Good Parts versus True Dimension

From Figure 3 it is clear that a good measurement system will protect you from misclassification until the true diameter approaches either specification limit, at which point there is a 50% probability of misclassification. Of course, once we consider diameters that are truly outside of the specification then the probability of failing a good part drops immediately to zero. Also, the probability of misclassification of good parts as failed is minimised when the true diameter is on target at 1.2500; a good reason why operators should aim for the target diameter!

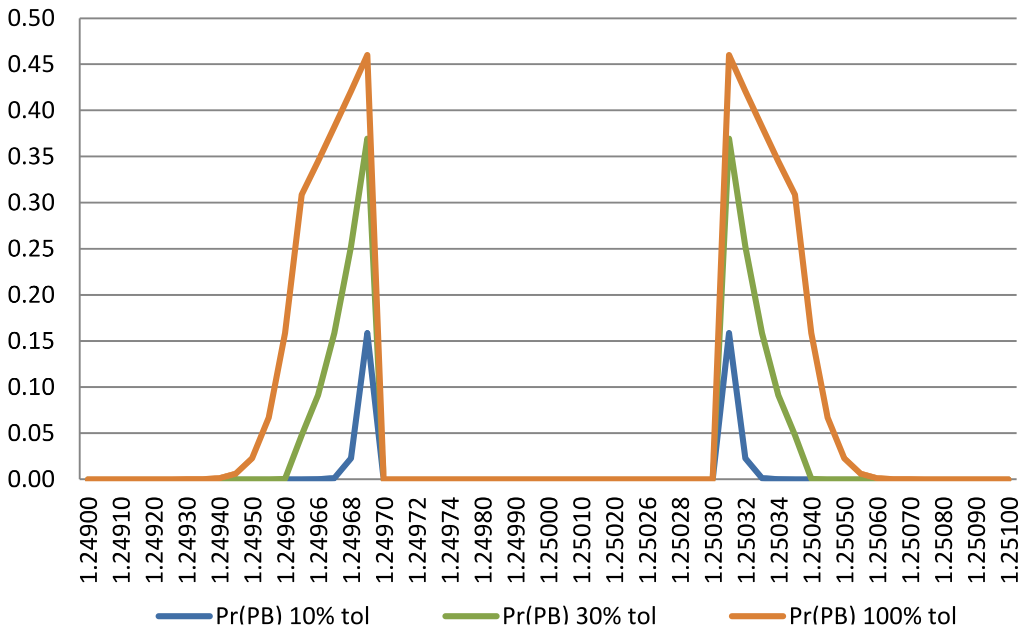

Figure 4 Probability of Passing Bad Parts versus True Bore Diameter

In Figure 4 we see that the probability of misclassification of bad parts is obviously zero for parts that are actually within specification. Also, the probability of misclassification for truly bad parts diminishes rapidly as the diameter gets further from the specification limit, and like the first case above, reaches a peak at the specification limit (a part which is on the specification limit is deemed good, so the maximum conditional probability of misclassification in this case is a little less than 50%).

If we consider the impact of *process capability* in conjunction with these results, it should be obvious that as process capability improves, we are less and less likely to make parts whose true dimensions approach or exceed the specification limits, and so the actual rates of misclassification will reduce accordingly. All in all these results highlight the benefits of: having a good measurement system, aiming for the target (nominal) dimension and having a good process capability (which will keep the part away from the dangerous specification limit regions).

5. Effect of Max. Metal Machining Practices on Product Performance

As has already been stated, MMM practices involve an incremental approach to producing components that are within specification limits: after an initial amount of material is removed, the component's dimension is measured. If it falls inside the specification limit, the component is accepted as conforming to the specification, otherwise an additional material removal pass and subsequent re-measurement is performed, and this process is repeated until the component can be accepted. It is fair so say however that there is no single, universal approach to the size and number of incremental material removal steps: two possible extremes of which would be to:

1. Plan to remove all material required to bring the part exactly on top of the aimed-for specification limit in a single material removal step, following which the part is measured to determine whether a further material removal pass (rework) is required. If rework is required, the amount of material to be removed next would be set to be the difference between the measured value and the specification limit.

2. Plan to remove the material required to bring the part exactly on top of the aimed for specification limit via a series of "small" finite steps, with measurements taken after each step in order to evaluate the progress made towards the specification limit and to adjust, if necessary, the next material removal step size as the specification limit is approached.

While it is certain that every different "style" of MMM that could be adopted in a particular circumstance may exhibit somewhat more or less of the issues that we will raise in this paper with regards to affecting product performance and to the likelihood of misclassification, it is also certain that each will be susceptible to both of these phenomena, and so for the purposes of brevity we have chosen to examine the first approach to MMM as outlined above.

The Bore and Shaft Max Metal Simulation

In order to further explore the impact of MMM it is necessary to also consider the *functional* impact of this practice on a product's performance. To illustrate this we will use the simple case of a bore and shaft assembled to an interference fit. As illustrated in Figure 5, The bore and shaft assembly is intended to achieve a nominal interference of 0.0010in, with allowable variation of ±0.0006in about this: if the interference is too low, the bore and shaft may lose either axial and/or circumferential location to each other under load; if the interference is too high, the assembly may be over stresses, which may lead to a diminished service life. In order to achieve the desired interference the bore and shaft have specified diameters of 1.2500±0.0003in and 1.2510±0.0003in respectively.

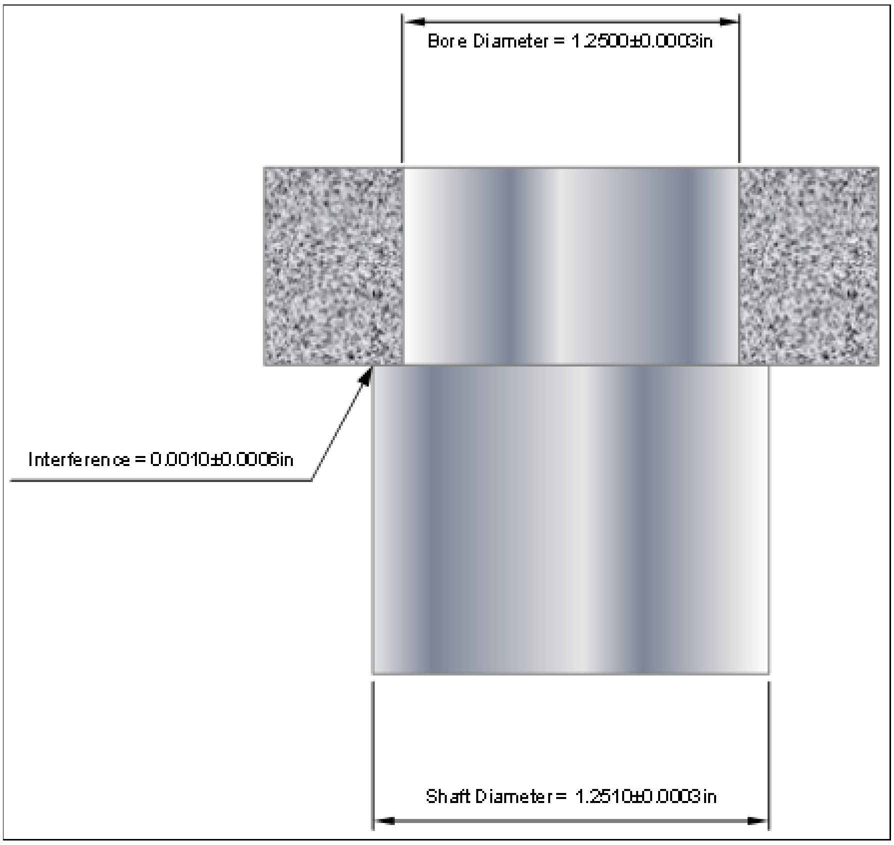

Figure 5 Bore and Shaft Assembly to an Interference Fit

The question now arises as to what the effect of a MMM approach to machining of these two parts will be in terms of not only the conformance of the parts to their individual specifications, but also performance of the expected assembly (as represented by how well the interference fit actually achieved conforms to specification).

Synthesis of Production Data of the Bore for MMM

To explain the assumed MMM process in operation we will first take the bore as an example. For the bore's inner diameter, parts that have diameters larger than the upper specification limit must be scrapped, whereas parts that fall outside the lower specification limit can be re worked, so under a MMM regime just achieving the lower specification limit would be aim.

In our chosen MMM scenario the operator aims to achieve the lower specification limit for the bore during each material removal pass, and after each pass the bore diameter will be measured. If the measurement indicates that the part is "bad" (still outside the lower specification limit), more metal will be removed; if the measurement indicates that the part is "good", no further material removal is considered necessary, and so the part is released and so passed on down through the latter stages of production.

Initially we will step through just one possible completion of such an iterative process, synthesising possible outcomes of each machining step and measurement taken in order to make the workings of the MMM scenario clear. After this, we will repeat the process multiple times in order to synthesise a possible production run for the bore, from which we can draw more general conclusions as to likely misclassification rates.

Note that in this MMM scenario we are assuming that when the operator sets the machine to turn an amount X from the diameter, it actually removes an amount according to a normal distribution centred on X and with a process standard deviation of σ_{pp} = 0.0000968in (inferred from assuming a %Study value of 25% determined from a Gauge R&R study). The inherent process capability is therefore C_{pk} = 1.055, which is "fair".

1. A bar is taken from stock (we will assume that the actual stock bore diameter is 1.2400in and does not vary – which must be an optimistic assumption!) and its diameter is measured by the operator as D_0

2. The operator will aim for a final measured diameter of 1.2497in, exactly on the lower specification limit by aiming to turn $T = D_0$ - 1.2497in.

3. The operator will then measure the new diameter as D_1. If D_1 is on or inside of specification, the part is passed as "good", otherwise, Step 2 is repeated, this time aiming to turn $T = D_1$ - 1.2497in.

For our purposes a value for %*Tol* of 25% will be used. This is not very good (i.e. it is not less than 10%), but such a value may in fact not be atypical for high precision applications (it may even be worse than this in practice). This %*Tol* value corresponds to a σ_{GRR} value of 0.0025in. The synthesis will use sampling from a normal distribution $N(0,\sigma^2_{GRR})$ for the measurement error and $N(T,\sigma^2_{pp})$ for the machining process, where T is the target reduction in diameter at each iteration.

Table 1 Record of Process for Production of One Bore Diameter

Step 0		
Actual Diameter	1.240000	outside spec.
Measured Diameter, D_0	1.240014	outside spec.
Action by operator	Remove metal	
Target Reduction in Diameter	0.009686	
Actual Reduction in Diameter (Monte Carlo)	0.009576	
Step 1		
Actual Diameter	1.249576	outside spec.
Measured Diameter, D_1	1.249601	outside spec.
Action by operator	Remove metal	
Target Reduction in Diameter	0.000099	
Actual Reduction in Diameter (Monte Carlo)	0.000027	
Step 2		
Actual Diameter	1.249603	outside spec.
Measured Diameter, D_2	1.249625	outside spec.
Action by operator	Remove metal	
Target Reduction in Diameter	0.000075	
Actual Reduction in Diameter (Monte Carlo)	0.000024	
Step 3		
Actual Diameter	1.249627	outside spec.
Measured Diameter, D_3	1.249593	outside spec.
Action by operator	Remove metal	
Target Reduction in Diameter	0.000107	
Actual Reduction in Diameter (Monte Carlo)	0.000057	
Step 4		
Actual Diameter	1.249684	outside spec.
Measured Diameter, D_4	1.249726	inside spec.
Action by operator	Release part	

In the specific case as detailed in Table 1, the final decision turned out to be *wrong*: a bad part was passed as good and the part went into stores with a recorded bore diameter of 1.249726in. Note too that even though it was intended to remove enough material to achieve the specification limit value of 1.2497in in a single pass, it actually required *four* machining operations to (apparently) achieve this due to the combination of measurement error and the inherent process variability.

As stated previously, we will now look at the outcome of synthesising a production run of 500 such bore diameters using this MMM scenario in order to determine possible misclassification rates. A sample of 500 diameters was synthesised by mimicking the process of Monte Carlo Sampling in Excel: a spreadsheet was created where individual rows represented each step and sub step as in Table 1, but with different (randomly sampled) outcomes each time. The final measured and actual diameters produced (i.e. once a measured diameter indicated that the part was "good") were then collated and presented as histograms, as shown in Figure 6.

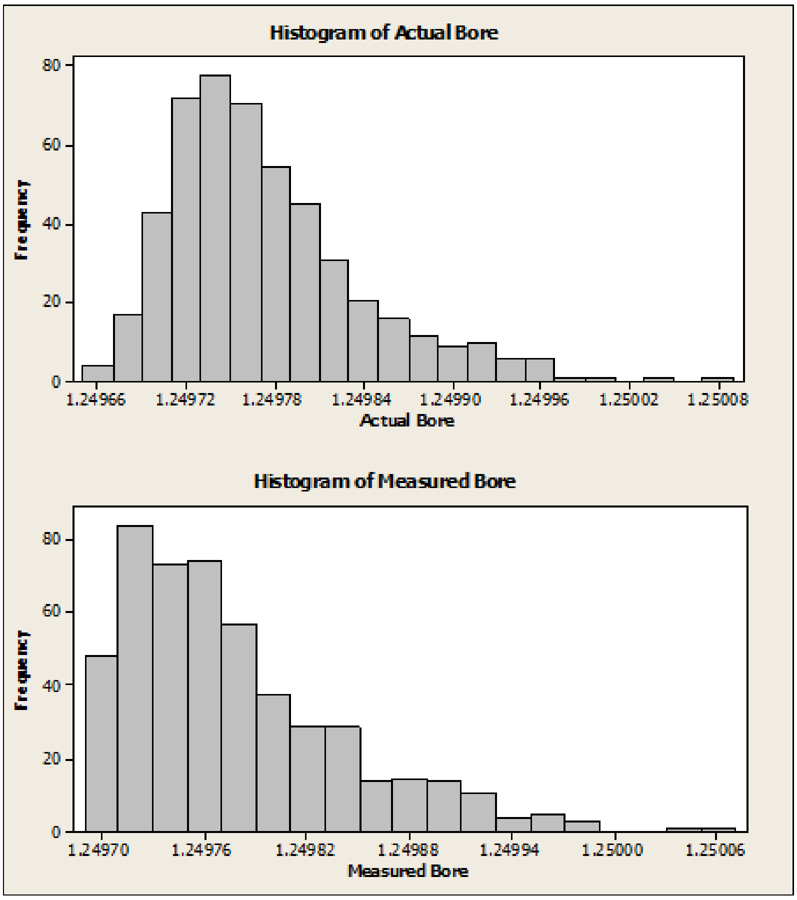

Figure 6 Comparison of Actual and Measured Bore Diameters for MMM

Considering first the histogram of measured bore diameters we can see that there is essentially a "cliff edge" quality to the distribution of the measured values against the lower specification limit of 1.2497in. In contrast, the histogram of actual bore diameters reveals that a proportion of diameters are outside of the specification. By count, there are 37 such instances out of the particular sample of 500, which equates to an observed misclassification rate for "passed bad" parts of 7.4%.

Given that this value is based solely on this single sample, and that another sample will inevitably have a different count of misclassifications, we need a more reliable means of estimating for the misclassification rate, which can be made by fitting an appropriate distribution to the sample and then using the properties of the distribution to predict the rate of misclassifications.

To do this we can use Minitab's *Individual Distribution Identification* facilities, which indicated that the distribution of actual bore diameters plausibly belonged to a *Largest Extreme Value* distribution with defining parameter values for location = 1.24974 and scale = 4.88898×10^{-5}, and so using these properties, Minitab's *Capability Analysis (non-normal)* estimated a misclassification rate of 8.15%, as shown in Figure 7 below. It is worth noting too that more than 99.9% of all bore diameters are smaller than the nominal value of 1.2500in as specified by design.

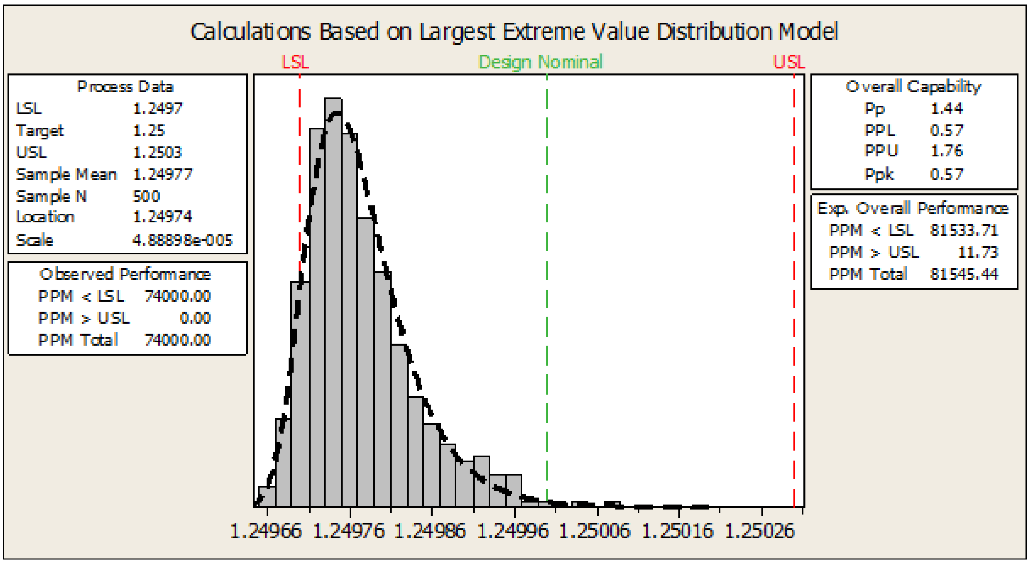

Figure 7 "Passed Bad" Misclassification Rates of Actual Bore Diameters for Max. Metal Machining

Synthesis of Production Data of the Shaft for MMM

In a similar fashion we can synthesise data for the shaft too. In the case of the shaft however, it is the *outer* diameter that is the critical dimension, and so parts that have diameters smaller than the lower specification limit must be scrapped, whereas parts that fall outside the upper specification limit can be re worked, so under a MMM regime just achieving the upper specification limit would be the aim.

Figure 8 shows the histograms for actual and measured shaft diameters, again for a sample of 500 shafts. Similarly to the observations made on the bore, we can see that again the histogram of measured shaft diameters again shows a "cliff edge" quality to the distribution of the measured values against the upper specification limit of 1.2513in. In contrast, the histogram of actual shaft diameters reveals that a proportion of diameters are outside of the specification. By count, there are 37 such instances out of the particular sample of 500, which equates to an observed misclassification rate for "passed bad" parts of 7.4% - a coincidentally similar figure to that of the bore.

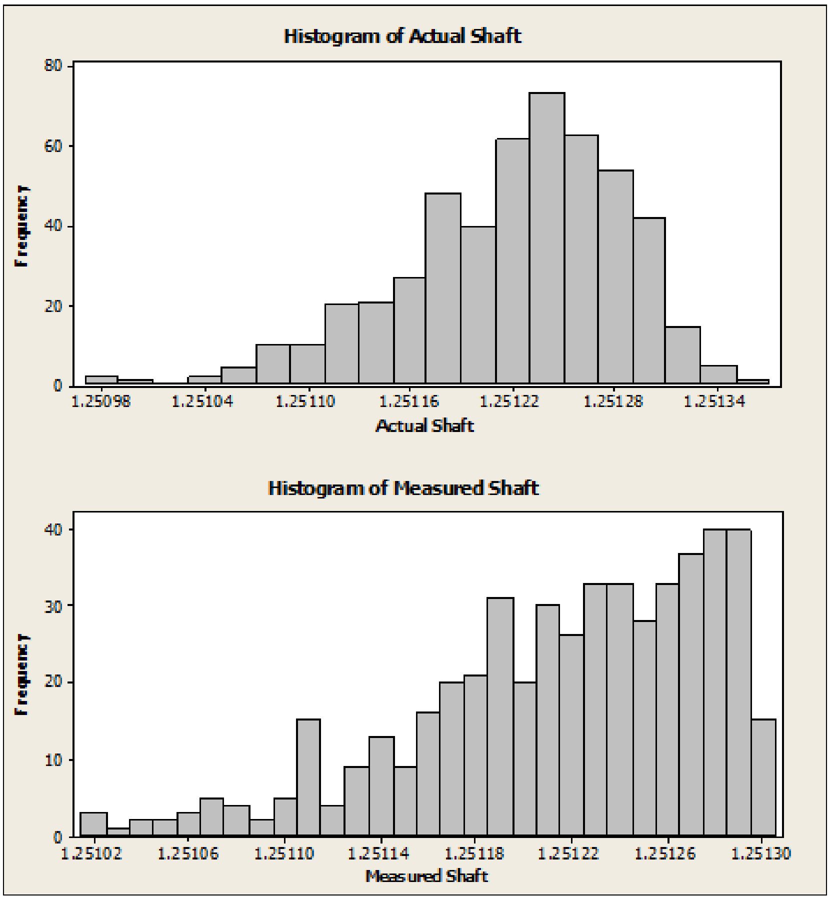

Figure 8 Comparison of Actual and Measured Shaft Diameters for Max. Metal Machining

As before, if we wish to have a more reliable estimate for the rate of misclassifications for the shaft, we need to fit a distribution and then use this to make our estimate. For the shaft, the most plausible distribution to fit our synthesised data was a *Smallest Extreme Value* distribution with defining parameter values for location = 1.25125 and scale = 5.16992×10^{-5}. Again, Minitab's Capability Analysis can be used to arrive at the estimate for misclassifications, as is shown in Figure 9, which indicates a rate of 7.33%. It is again worth noting too that more than 99.9% of all shaft diameters are larger than the nominal value of 1.2510in as specified by design.

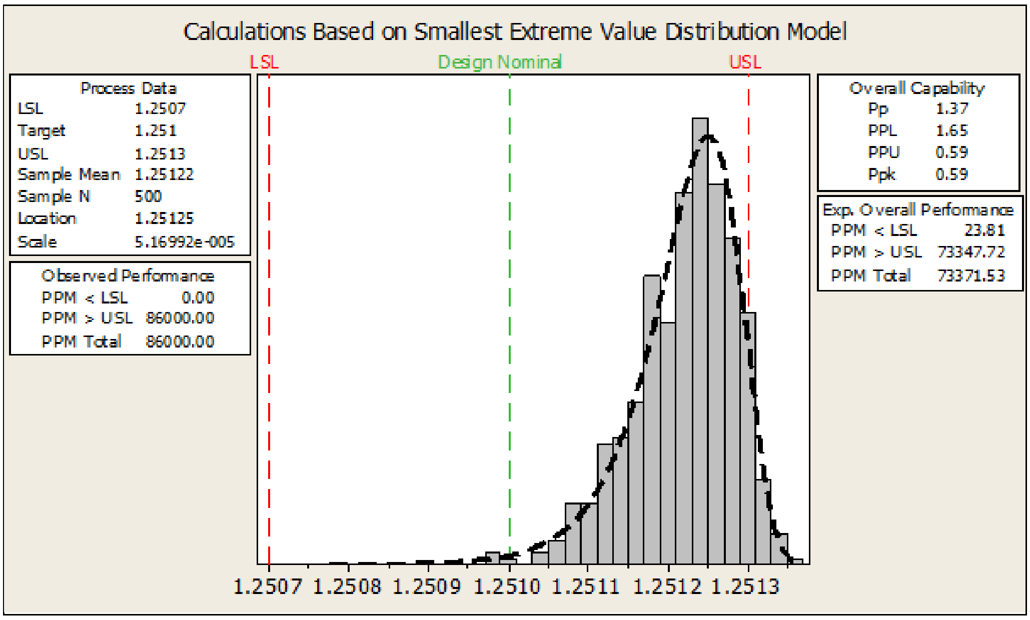

Figure 9 "Passed Bad" Misclassification of Actual Shaft Diameters for Max. Metal Machining

Variation in Interference Fit of the Bore-Shaft Assembly for MMM

Finally, we need to consider how the distributions of bore and shaft diameters have affected the interference fit of the assembly. As stated earlier, the achievement of an appropriate interference fit in the specified range is important to the function and longevity of the assembly – and ultimately to the product of which this is a perhaps key part.

A similar sample of 500 interference fits was generated by assuming that each bore and shaft would be assembled as a pair as they each came off the production line, and so the actual interference fit was taken to be the difference between the actual bore and shaft diameters. For the purposes of comparison, an "assumed" interference fit for each pair was also derived, being the difference between the measured values of bore and shaft diameters. The resultant histograms are shown in Figure 10.

In this context however, we do not interpret the interference fits in terms of misclassification, rather in terms of defect rate compared to the performance specification of 0.0010±0.0006in. In this context the assumed interference does not indicate any defects (albeit given that there is no equivalent "cliff edge" to the assumed interference fits there would seem to be a possibility of such), whereas the actual interference fits, being somewhat more biased towards higher values of interference, and in fact by count there were 6 defective assemblies with interference fits above the upper limit of 0.0016in.

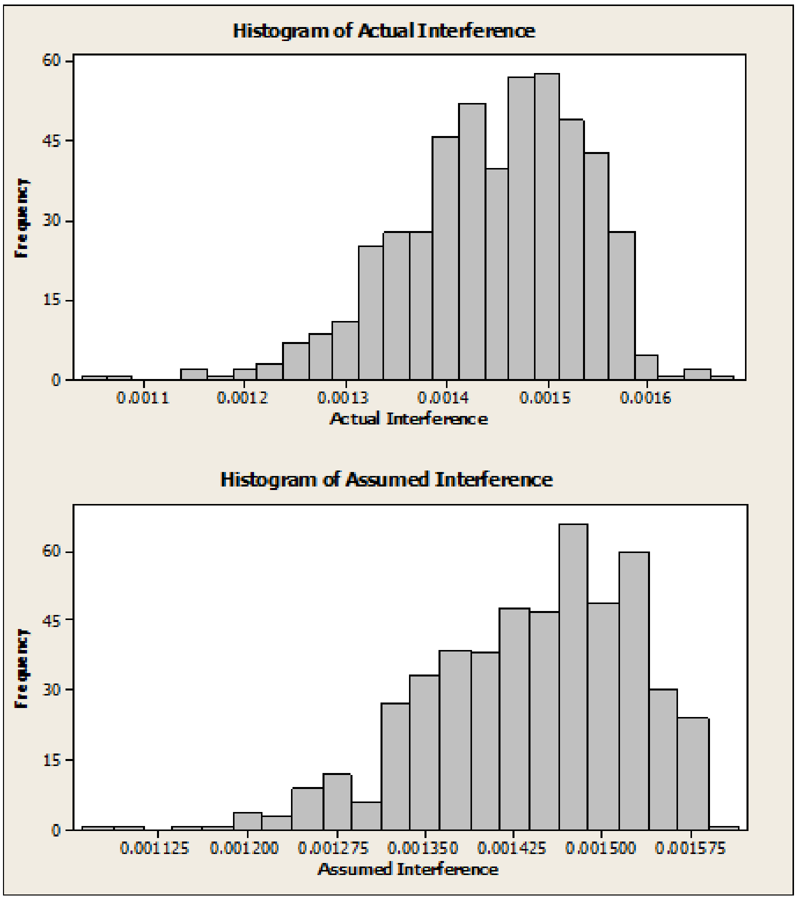

Figure 10 Comparison of Actual and Assumed Interference Fits for Max. Metal Machining

Analogous to the case for both diameters, if we wish to have a more reliable estimate for the rate of defects for interference fit, we again need to fit a distribution to the synthesised data and then use this to make our estimate. For interference fit, the most plausible distribution to fit our synthesised data was a *Weibull* distribution with defining parameter values for shape = 19.0515 and scale = 0.00148764. The resulting Capability Chart is shown in Figure 11, indicating an expected defect rate of 1.83%. It is again worth noting too that more than 99.9% of all interference fits are higher than the nominal value of 0.0010in as specified by design, which may mean that typical life of the assembly is consequently lower than predicted.

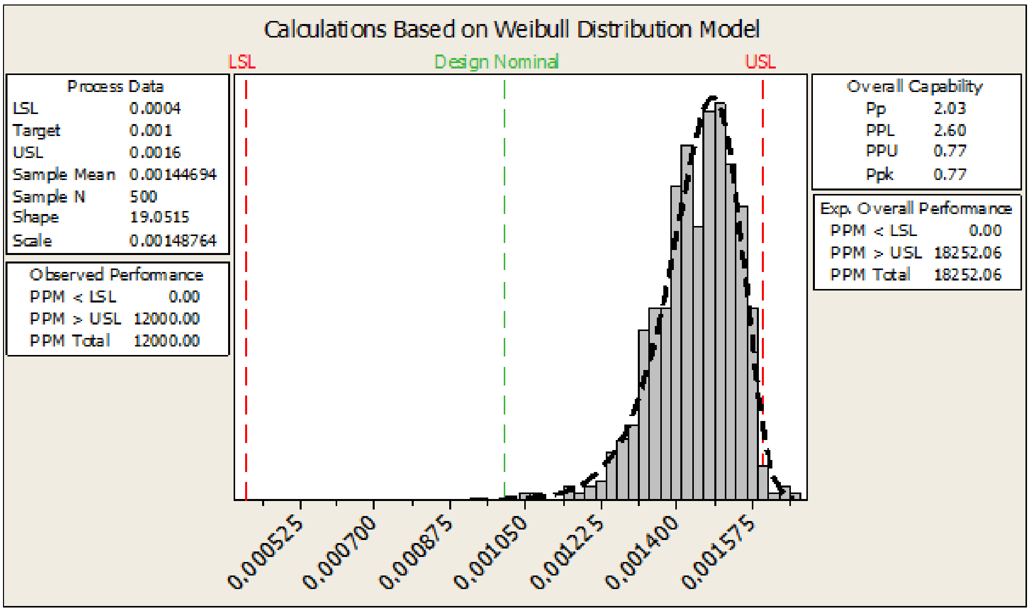

Figure 11 *Process Capability of Actual Interference Fits for Max. Metal Machining*

6. Application of On Target with Minimal Variation Practices on Product Performance

In comparison to the results for the MMM approach to producing bore and shaft assemblies we will now consider the alternative case of applying *On Target with Minimal Variation* (OTWMV) principles, using exactly the same design specifications and levels of variation present in both the measurement system and the underlying material removal process that were used for MMM. The essential and significant difference is that instead of targeting the lower or upper specification limits for the bore or shaft respectively, both parts will have the design nominal value as the target instead.

Synthesis of Production Data of the Bore for OTWMV

The resulting histograms for the bore diameter are as shown in Figure 12. Here we can see that (as might be expected, given that both the measurement system and the underlying process were assumed to be normal), the distribution of both actual and measured bore diameters are plausibly normal (this is confirmed using Minitab's *Individual Distribution Identification* process); the difference between actual and measured here is simply the "inflation" of the observed part-to-part variation due to the measurement error, i.e. $\sigma^2_{tot} = \sigma^2_{pp} + \sigma^2_{GRR}$. It should be noted that while there were no observed defects generated in this sample, there is at least a small possibility of doing so.

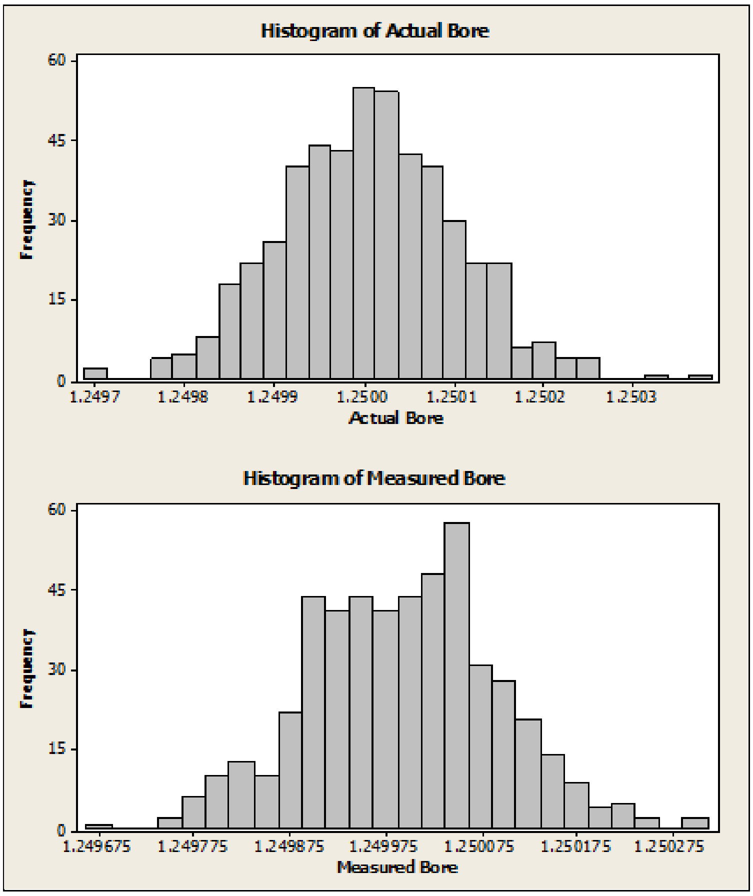

Figure 12 Comparison of Actual and Measured Bore Diameters for On Target with Minimal Variation

Once again, performing a *Capability Analysis* on bore diameter (this time on both actual and measured) will give us an indication of the underlying process capability and misclassification rates. Comparing Figure 13 and Figure 14, the actual defect rate for the process is predicted to be 0.23%, whereas the presumed defect rate (based on measured parts) is marginally worse at 0.26%. Intuitively this indicates that the rate of misclassification for bore diameters is in this case negligible.

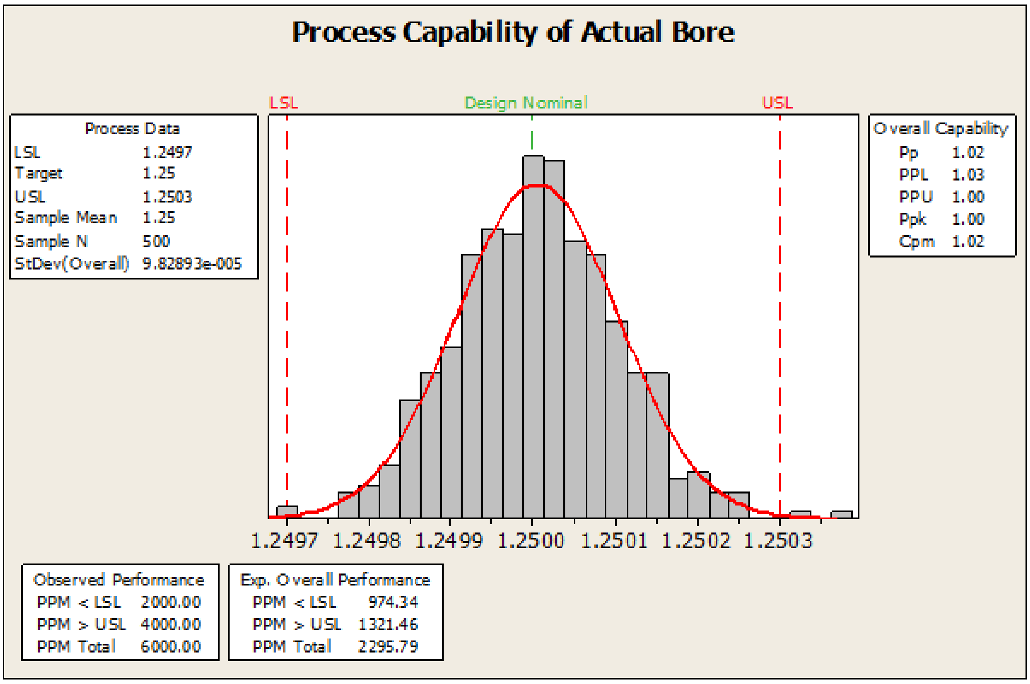

Figure 13 Process Capability of Actual Bore Diameter for On Target with Minimal Variation

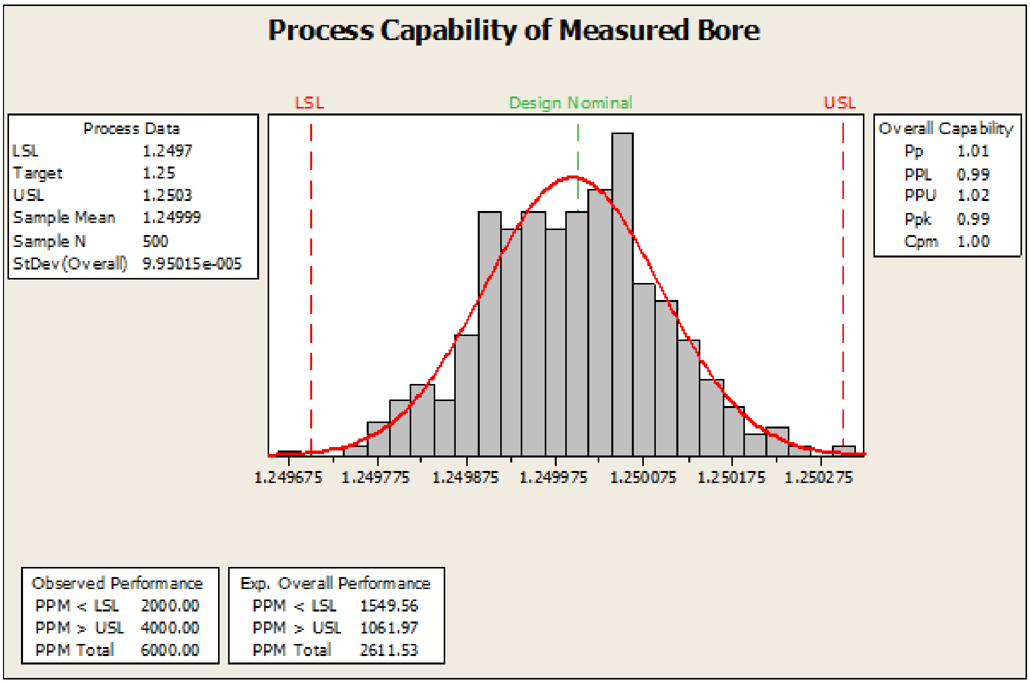

Figure 14 Process Capability of Measured Bore Diameter for On Target with Minimal Variation

Synthesis of Production Data of the Shaft for OTWMV

If we now look at the equivalent charts for the shaft diameter we see a very similar picture: a comparison of both actual and measured shaft diameters shows that they are both plausibly normal (this is again confirmed using *Minitab's Individual Distribution Identification* process) and that the difference between actual and measured here is simply the "inflation" of the observed part-to-part variation due to the measurement error. It should once again be noted that while there were no observed defects generated in this sample, there is at least a small possibility of doing so.

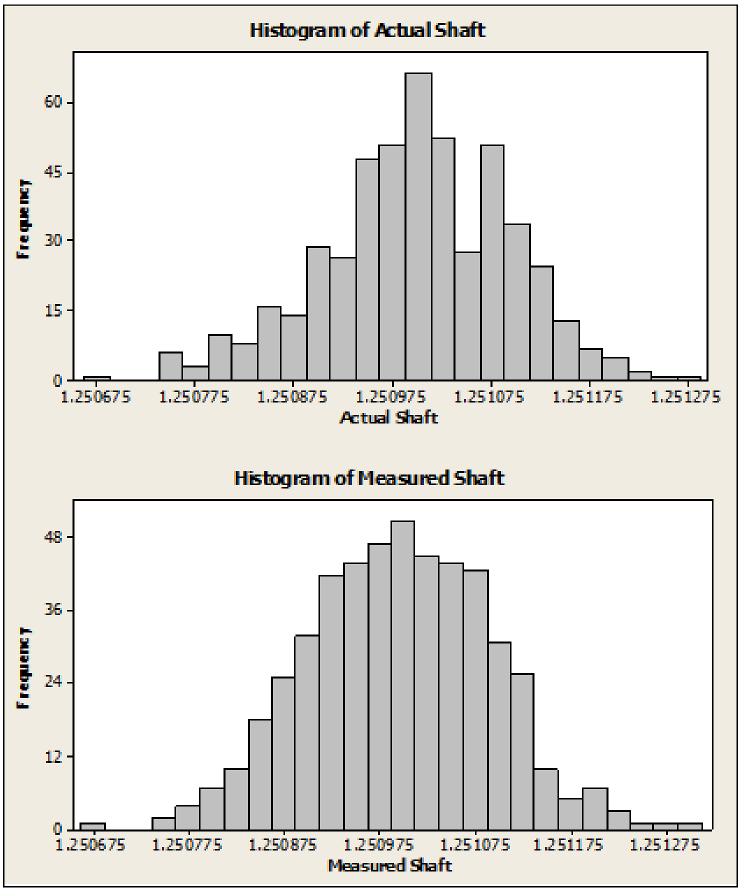

Figure 15 Comparison of Actual and Measured Shaft Diameters for On Target with Minimal Variation

Once again, performing a *Capability Analysis* for both actual and measured shaft diameter gives us an indication of the underlying process capability and misclassification rates. Comparing Figure 16 and Figure 17, the actual defect rate for the process is predicted to be 1.54%, whereas the presumed defect rate (based on measured parts) is marginally worse at 1.83%. Once again, the rate of misclassification for shaft diameters is negligible.

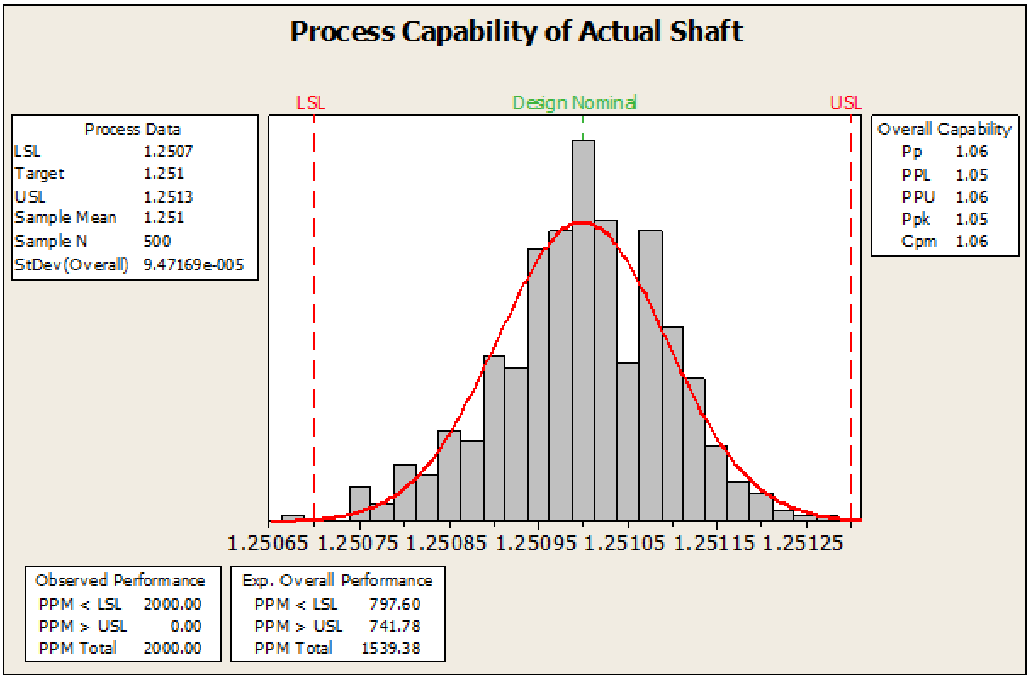

Figure 16 Process Capability of Actual Shaft Diameter for On Target with Minimal Variation

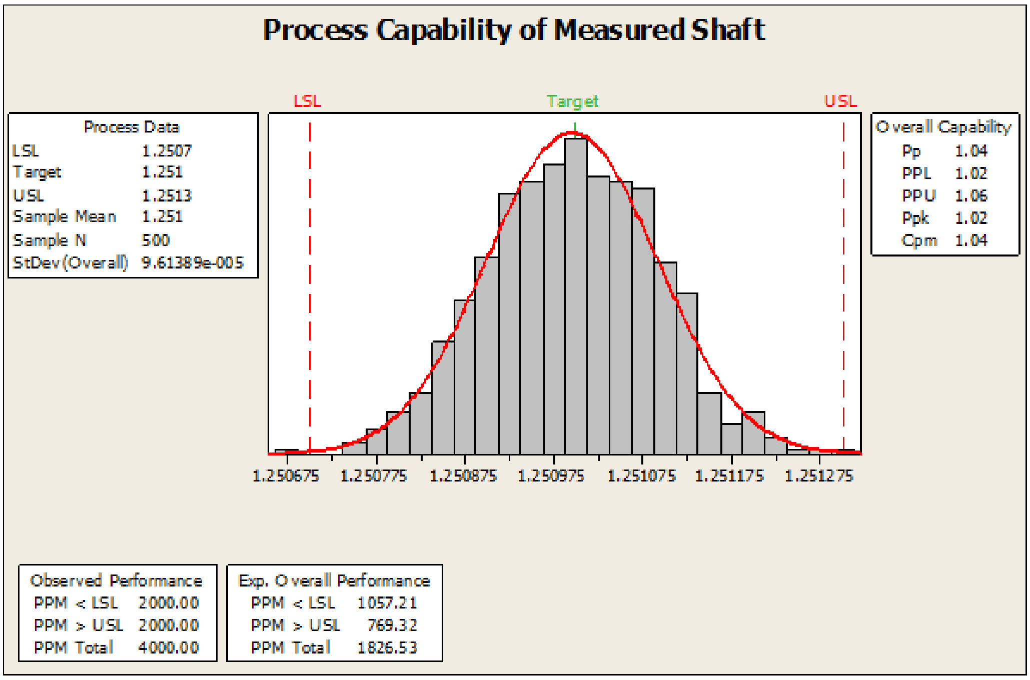

Figure 17 Process Capability of Measured Shaft Diameter for On Target with Minimal Variation

Variation in Interference Fit of the Bore-Shaft Assembly for OTWMV

Finally, we of course need to consider how these distributions of bore and shaft diameters have affected the interference fit of the assembly, since the achievement of an appropriate interference fit in the specified range is important to the function and longevity of the assembly – and ultimately to the product of which this is a perhaps key part.

As for the MMM case, a sample of 500 interference fits was generated by assuming that each bore and shaft would be assembled as a pair as they each came off the production line, and so the actual interference fit was taken to be the difference between the actual bore and shaft diameters. Again, for the purposes of comparison, an "assumed" interference fit for each pair was also derived, being the difference between the measured values of bore and shaft diameters. The resultant histograms are shown in Figure 18.

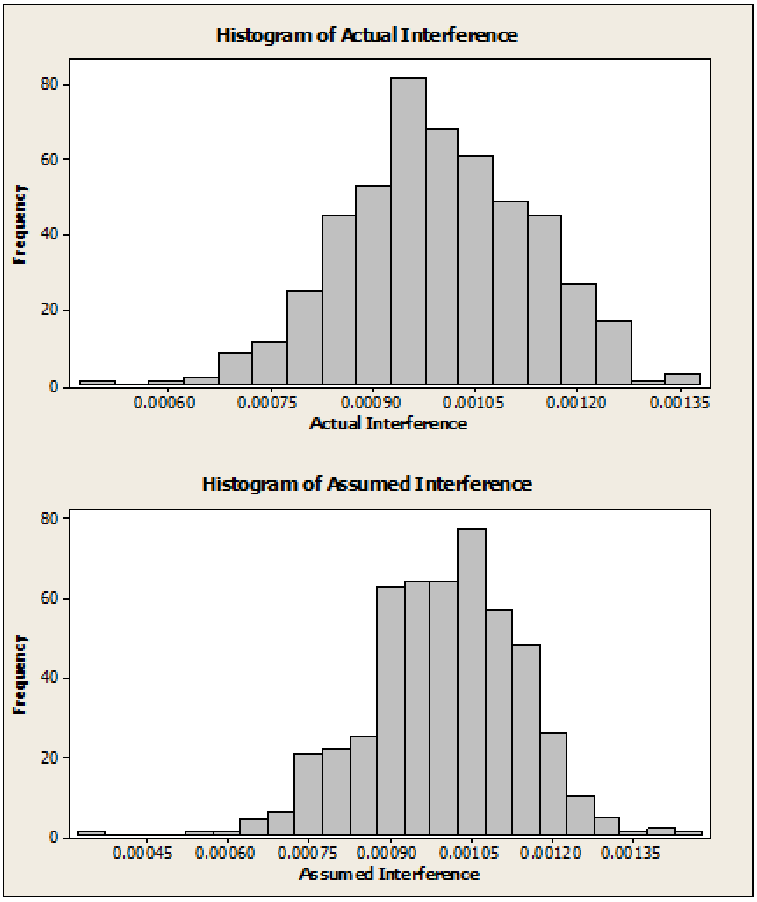

Figure 18 Comparison of Actual and Assumed Interference Fits for On Target with Minimal Variation

Just as for both bore and shaft diameters, the actual and assumed interference fits are plausibly normal (again confirmed using Minitab's *Individual Distribution Identification* process) and also exhibit the same slight "inflation" in the assumed interference fit's distribution in comparison to the distribution of actual interference fits. It is interesting to note that even visually, there seems to be only a small probability that the assembly would produce an interference fit that falls outside either the lower or upper specification limits of 0.0004in and 0.0016in respectively. To support this presumption, we need of course to examine the results of the *Capability Analysis* for interference fit.

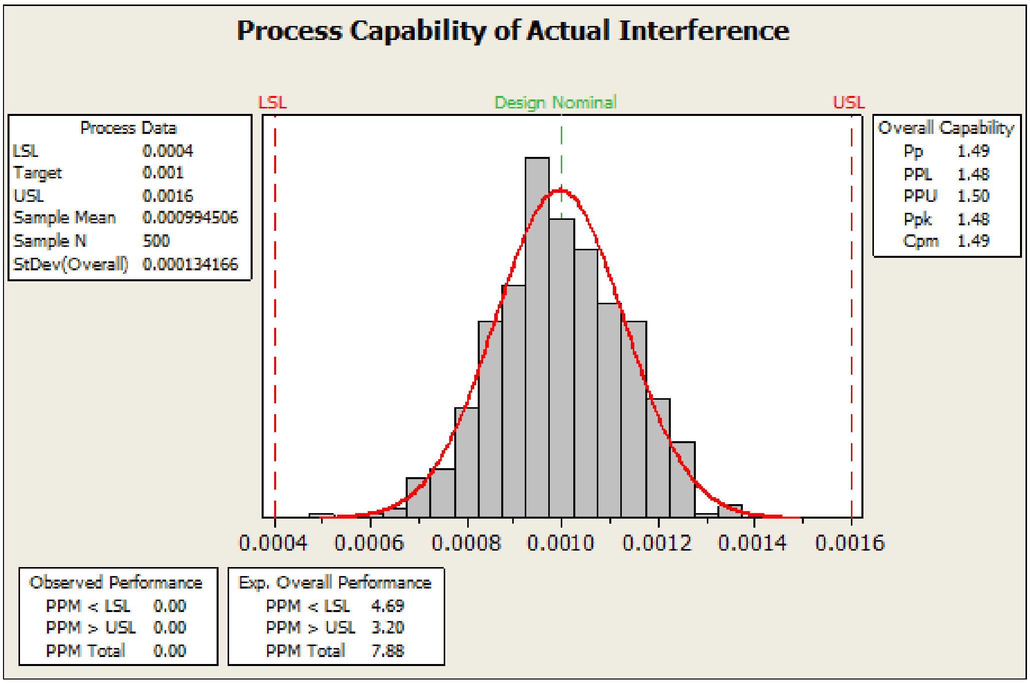

Figure 19 Process Capability of Actual Interference Fits for On Target with Minimal Variation

As we can see in Figure 19, the predicted total defect rate for interference fit is estimated to be just 0.0008% and the mean interference fit achieved is on target at 0.0010in. It is noteworthy too that the process capability achieved for interference, which links strongly to the ability for the assembly to perform its required functions (maintain axial and circumferential co-location) and achieve the required service life, is a C_{pk} of 1.49. Note that Minitab labels the process capability as P_{pk} since it assumes the data provided also includes long-term sources of variation rather than just short-term.

If we consider the C_{pk} achieved for interference in relation to the current values for the bore and shaft diameters (a C_{pk} of 1.0 and 1.05 respectively as reported in Figure 14 and Figure 16), Then we should perhaps consider whether the current specification limits for the bore and the shaft diameters are "appropriate". If these limits were relaxed, then *as long as the current process output were to remain unchanged*, even the current small rate of "defects" can essentially be eliminated without affecting the performance characteristics that ultimately we are most interested in: the assembly's interference.

7. Conclusions

This paper has explained the theoretical grounding for misclassification of measured parts when compared to a specification. This grounding was based upon the cumulative distribution function of a joint (bivariate normal) distribution for the underlying parts and their measured (observed) values.

Via a simple example, the results of a *Monte Carlo Simulation* using the same parameters were shown to agree well with these calculations (as expected). The advantage of using MCS is that it can accommodate many different distributional assumptions for both the underlying process of actual dimensions and the random error distribution. It is also easy to introduce bias of the measurement system into the simulation by simply changing the mean error (or location parameter in general) from zero to whatever is required.

We have also shown, through synthesising data for a typical bore and shaft assembly, that the consequences of adopting an approach whereby the process is operated close to either specification limit, such as in the common practice of *Max. Metal Machining*, is that the potential for misclassification of parts is greatly increased. Since both the likelihood of parts being "passed bad" or "failed good" is increased, there will be an increase in the level of re-work perceived necessary during the machining process, leading to increased cost (a consequence of the "failed good" rate), as well as the inevitable level of actually defective parts produced (a consequence of the "passed bad" rate).

Furthermore, such practices have been shown (through the same synthesised example) to have potentially significant ramifications for the ability of a design (of which a part is just one component) to fulfil its functional (and hence customer) requirements: even though all parts are measured) as conforming, the bias in part dimensions away from the design nominal, coupled to the presence of actually defective parts mean that the product itself will exhibit significant levels of non-conformance too.

While the possibility of misclassification is not restricted to *Max. Metal Machining*, any process that operates near one or both specification limits is more susceptible to misclassification than is a process that targets the middle of the specification range (i.e. the design nominal). It has been shown (once again using the same synthesised example) that the *On Target with Minimal Variation* approach is a viable alternative, as not only will the rates of misclassification be lower, but also the product performance will be likely to be much higher conforming too.

Finally, it has been shown that through linking the statistics of the underlying process variation in individual parts to the required variation in functional performance (in the synthesised example used, this being represented through a required level of interference fit), we can determine the most appropriate settings for the part's specification limits. Such statistically derived specifications (*Statistical Specifications*[4] for brevity) essentially allow us to agree an appropriate level of process capability to be met such that the required rates for both scrap and re-work will be met, assuming the specifications are not used as "goal posts" and the process mean and variability are maintained going forwards. Processes that are specified in such a way are also not susceptible to misclassification either, by definition.

8. References

1 Minitab® 17
Minitab Inc., Quality Plaza, 1829 Pine Hall Road, State College PA 16801 3210, USA
http://www.minitab.com/

2 Oracle Crystal Ball
Oracle Corporation, 500 Oracle Parkway, Redwood Shores,
CA 94065, USA
http://www.oracle.com/us/products/applications/crystalball/overview/index.html

3 Microsoft Excel
Microsoft Northern California District Office, 1065 La Avenida, Mountain View, CA 94043
http://products.office.com/en-GB/excel

4 Rowe, P. G.
Statistical Specifications for Critical to Quality Characteristics,
International Journal of Six Sigma and Competitive Advantage,
2006 Vol. 2, No. 1 *pp*89-103.