Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

The importance of being repeatable

T. J. Morris
Cincinnati Machine UK Ltd, PO Box 505,Kingsbury Rd, Birmingham, England B24 0QU, Email: tim_morris@cinmach.co.uk

Abstract

The introduction of the latest (1997) version of ISO 230-2—the ISO Standard for Repeatability and Accuracy of Positioning of Machine Tools—has changed the way we calculate Repeatability. 'Standard Uncertainty' has replaced 'Standard Deviation', and the band width has been reduced.

This new concept of Repeatability is examined from both a theoretical and a practical standpoint, and evidence is presented of the way results vary and what sorts of distributions of Repeatability occur in practice.

The scope of accuracy measurements and the time scale are then expanded, since positioning is not the only factor controlling a machine's overall accuracy: thermal problems, in particular, are discussed.

Problems of maintaining Repeatability in the longer term are discussed with particular reference to installing the machine in a customer's premises.

1 The New Version of ISO 230-2

The publication in 1997 of the new version of the ISO standard ISO 230-2[1] for Repeatability and Accuracy has forced us to look again at what we mean by Repeatability. The old standard of 1988 had much in common with most of the other existing standards at that time: BS3800/2 (originally BS4656/16), VDI/DGQ 3441 and the old NMTBA procedure from around 1970. All these embraced the concept of Repeatability being measured as a six standard deviation spread on an assumed Normal distribution curve. Putting it very simply, this meant that we believed that nature had carefully arranged things so that we could, by measuring a few results, compute the overall Repeatability that would be expected (with 99.7% confidence) from a much larger set of data.

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

Opinions have long been divided as to whether this was a valid assumption, and to be fair, most people, if asked to evaluate Repeatability, would intuitively take a handful of readings and measure the total spread. The new Standard now concedes that the distribution may not in fact be Normal after all, and side-steps this difficulty by changing **Standard Deviation** into **Standard Uncertainty**—but we should note that Standard Uncertainty is calculated in exactly the same way as Standard Deviation. At the same time, it reduces the acceptable band-width from six to four. Using four instead of six lowers the confidence level from 99.7% to 95%, and lines up with current ISO practice. The statistical approach to measuring Repeatability (and hence Accuracy) finds parallels in Statistical Process Control used in manufacturing engineering.

Such a change was bound to create confusion. Manufacturers of Machine Tools publish specifications like 'Accuracy and Repeatability to ISO 230-2' in their glossy brochures. Now they find that Repeatability values in the 1997 version are magically 'improved' by 33% over the previous values (1988). Clearly, we must now be very diligent when quoting against this Standard, for omission of the year will make nonsense of the values. (An interesting repercussion from this is that the DTI are changing their accuracy criteria used in assessing the control of export licences for machine tools.)

Opposing this change on the grounds of mere pragmatism was bound to fall on deaf ears; and now that some of the dust has settled it seems that there might indeed be a benefit to using this new approach. But first, we need to look a little more closely at what we mean by Repeatability, and why it exists as a concept at all. (The significance of the other measurement parameter 'Accuracy' was covered in an earlier Lamdamap paper.[2])

2 Repeatability

2.1 The Normal Distribution

We know that if we make repeated attempts to position a machine axis at a particular target point we get a spread of results that increases quite quickly at first, but then slows down, growing at an ever decreasing rate. If the results follow our assumed 'Normal Distribution' we can then calculate the standard deviation and note that a fixed percentage of results will lie at certain distances from the mean. Of particular interest is the fact that 68.5% lie within one standard deviation, 95.4% within two, and 99.74% within three of the mean.

The Normal distribution also tells us that however far we stray from the mean, we can never quite get to a point where 100% confidence is achieved, i.e., where all possible results are embraced. There is always a remote chance of getting some rogue value: at seven standard deviations from the mean this chance is 2 in 10^{12}! Experience, however, persuades one to believe that physical constraints really do put finite limits on the distribution.

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

Since our traditional plus or minus three standard deviations gives us a theoretical 99.74% confidence, in a sample of one thousand readings we should expect to get all but two or three values within this range. Standard deviation is a statistical property of an infinite population; what we obtain from our test results is **Dispersion** which is an estimate based on our much smaller sample. (Note that dispersion uses the symbol s, whereas standard deviation uses σ.) Expressed in dispersion bandwidths, we previously defined our Repeatability to be **6*s***. Now it has become **4*s***, and this gives us our 95.4% confidence—but only when the distribution is Normal.

This assumption (of Normal distribution) has now been dropped from the ISO standard by employing the term **Standard Uncertainty** for the measured s. Whether or not positioning results actually follow a Normal Distribution is another matter: there is no *a priori* reason for supposing this. Research work has shown that some machines do exhibit distributions approximating to this shape, when thermal fluctuations can be contained. This is not accepted universally and it would seem that, as we guessed, physical limits to the extent of the distribution really do exist. Clearly, the situation varies from machine to machine and even on the same machine. (See, for example, figures 1 and 2.)

2.2 Other Distributions

Other distributions may be postulated. For instance, if results are randomly distributed between finite limits that are determined by the system resolution we should get a **Rectangular Distribution**; if the machine is subject to a varying temperature, through some form of thermal control, then a **Sinusoidal Distribution** might be expected.

Although σ is a parameter of the Normal distribution function, this does not mean that its *alter ego*, s, cannot be calculated, or have any meaning, for these other types of distribution: it still remains as a measure of how far values stray from the mean. We have seen that if a Normal distribution exists then the relationship between s (or σ) and the confidence value is defined mathematically and may be obtained from statistical tables. For distributions with finite limits the confidence must always be 100% and the overall **Range** can then be related to a measured s for a large sample. The ranges of such sinusoidal, rectangular and normal distributions (for large sample sizes) are shown in table 1. The introduction by ISO of the **4*s*** range (a 'coverage factor' of two) is thus probably more realistic in that it is fairly close to the range for the rectangular distribution. It is clearly much closer to these more practical distributions than the old **6*s*** can be.

What we are saying here is that when we do a test and calculate **4*s*** then the 'true' range will depend on the distribution, the shape of which is not generally known. If happens to be a Normal distribution then **4*s*** gives us 95.4% of results; if it is rectangular then **4*s*** over-estimates the range by 15% (4/3.46).

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

Table 1. Range as a function of *s* for different distributions

Type of distribution	Range
Sinusoidal distribution	2.83*s*
Rectangular distribution	3.46*s*
Normal distribution at 95.4% confidence	4*s*
Normal distribution at 99.7% confidence	6*s*

2.3 Range *v* Sample Size

It is a property of a Normal distribution that the mean range of a given sample size can be used for computing *s* (and hence σ). Statistical tables are available giving σ against **Range** for up to 10 samples. This is an acceptable, though nowadays unnecessary, alternative to calculating it by formula. Some older standards offered this method as an aid to calculating *s* and provided a table of the required values. Much more interesting than this is the fact that it can be use to predict the range of a sample in terms of so many standard deviations.

From such a table, we can see for instance, that the mean range for a sample size of four is given as **2.059*s*** and for nine it is **2.97*s***. Using a computer, we can extend this table for larger samples and show that to get a range of **4*s*** we would need a sample of **27,** and to get **6*s*** would require approximately **444**.

This is an important concept: it means that with a Normal distribution the *calculation* of **4*s*** or **6*s*** from just a few results should give the same result as *measuring* the overall range from 27 or 444 readings respectively!

Using the range from a sample to calculate *s* is not the same thing as using the range as a measure of dispersion *instead* of *s*. But what we can do is to measure the range for one sample size and use it to compute the expected range for a much larger sample size. For example, suppose that the measured range from a sample of nine (i.e. nine approaches to a single target) is **0.010mm**, we know that for a sample of 9 this range should be **2.97*s***. So s = 0.010/2.97 = 0.0034. Our **4*s*** range dispersion is 0.0034 x 4 = **0.0136mm**. (The **6*s*** dispersion is 0.0034 x 6 = **0.0204mm**.) So, from a sample size of 9 we have derived the **4*s*** figure, which is the same as would be obtained from measuring the range directly from 27 approaches. The **6*s*** figure would similarly be obtained from 444 approaches. For a Normal distribution, the sample size should not be important. A larger sample should increase the accuracy of the results (though not necessarily the accuracy of the machine!) by making the distribution curve closer to the ideal—i.e. more of a curve and less of a histogram. On the other hand, larger samples do require a longer time to collect and the longer time provides more opportunity for external disturbing influences to occur.

2.4 Range Method for Repeatability

Other standards, notably from Japan (JIS)[3] and the USA (ANSI/ASME)[4] eschew the statistical approach altogether and prefer to employ the simple range

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

method using a fixed number of approaches. Here, the sample size is more likely to be significant. We may compare dispersion values against the ranges obtained by these different sample sizes for our Normal distribution.

Assume that the **4*s*** Repeatability value for an ISO test gives a value '**R**', then table 2 shows what values of Repeatability would be expected from the same machine using these different test methods. These are what we get by applying non-statistical test methods to a Normal distribution. If we now compare the results for the other types of distributions shown in table 1 and 'normalise' them to fit in with table 2 we get table 3.

Table 2. Repeatability from different test methods

Test method	**Repeatability**
ISO test using 4s (1997)	1.0R
ISO test using 6s (1988)	1.5R
JIS method: range over 7 approaches	0.7R
ANSI method: range over 10 approaches	0.8R

Table 3. Repeatability for different types of distribution

Distribution	**Repeatability**
ISO test using 4s (1997) (= range over 27)	1.0R
Rectangular distribution: full range	0.9R
Sinusoidal distribution: full range	0.7R

Thus the value obtained for Repeatability is dependent upon the method chosen to evaluate it. Which should we use? With the simple 'range' method we take a chance: the more readings we take, the more likely the chances of trapping extreme values, with the results getting inexorably 'worse'. With the statistical approach we get nearer the 'truth' by taking more readings, and the results get 'better' as the distribution approaches the ideal shape. But, as we have seen, it becomes increasingly difficult to maintain stable test conditions.

Experience shows that distributions may be Normal up to a band width of about **4*s***, but after that the distribution drops off quite sharply, approximating to something nearer the rectangular shape. Looking at the values in tables 2 and 3, we can see that the **4*s*** value chosen by ISO always slightly over-estimates the values obtained by other methods and by other distributions. Fortuitous this may be, but the current metrological practice of using a **4*s*** band would seem to give us a good working approximation to what we might expect in the real world.

2.5 Practical Tests to Study Distribution

Tests were carried out on a machine that had reached thermal stability by approaching a series of four targets 25 times each. Each test was then repeated a

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

further three times, making a hundred approaches in all. (Pairs of targets were at the same location but approached from opposite directions.) Statistical tables predict that for a sample size of 25 and a Normal distribution we should get a range of **3.931*s***. The first two targets showed a mean range of **3.8*s*** and the second pair a mean range of **2.8*s***. Put another way, the measured range over 25 approaches was 80% (+/-20%) of the predicted **4*s*** range. The second pair of targets generally had a larger variation because of increased mechanical and thermal compliance at the point of measurement. So when the Repeatability was good the **4*s*** value made quite a good estimate; when the Repeatability was poorer (i.e. larger) the **4*s*** value over-estimated the results by about 20%. Quite clearly, the **6*s*** values always grossly over-estimated the result. When the sample size was doubled the increase in range varied according to the degree of thermal drift present. The tables predict that the range should increase by 14% which was about right, on average.

When we plot out the results on a histogram, some are remarkably close to the ideal 'normal' curve, while others are nothing like it at all. Our two pairs of targets 1&2 and 3&4 illustrate this graphically—see figures 1 and 2. We are thus presented simultaneously with arguments both for and against the acceptability of a Normal distribution.

2.6 Correlation Coefficients and Thermal Drift

Another important fact that emerged from this test was that despite the machine being thermally stable—it had actually run the ISO230-3 thermal test first—it was still not possible to rule out thermal drift *absolutely*. The existence of thermal drift can be established by measuring the **Correlation Coefficient** of positional error with time. (Any statistical package that produces standard deviation can usually also generate the correlation coefficient.) Over a short period, such as that required to carry out an ISO test on a machine, thermal drift will vary reasonably linearly with time, and this fact can be exploited to detect its degree of prominence. The correlation coefficient is a number between -100% and +100%. Values near zero indicate little correlation and hence little thermal drift; while the nearer they approach unity the more probable becomes thermal drift. Even when great care has been taken to achieve constant temperature conditions, it is still not uncommon to get values of 80% or more which indicate that thermal drift still prevails. The sign of the coefficient indicates whether warming or cooling is taking place.

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

Figure 1. Distribution for targets 1 & 2

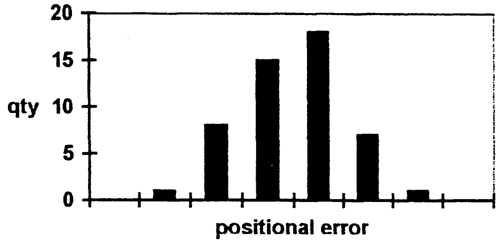
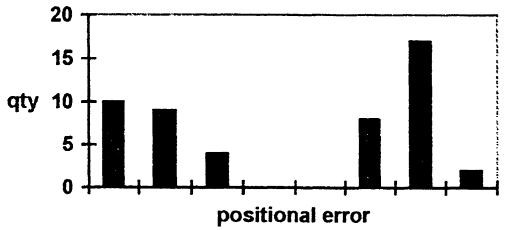


Figure 2. Distribution for targets 3 & 4



In our test, targets 1 and 2 started off with a correlation coefficient of 87-90% (indicating the strong presence of thermal drift), and gradually reduced to 65%. The other pair, despite—or maybe because of—being more compliant had correlations reducing from 40% to 20% (which indicated much less linear thermal drift). The gradual reduction in thermal drift showed that the machine was stabilising to a new set of operating conditions. Surprisingly, the presence of drift does not necessarily distort the Normal distribution too much. For example, drift applied to an otherwise rectangular distribution will turn it into a triangular one which is somewhat closer in appearance to a Normal curve.

Because thermal drift can be quantified through the measure of the correlation coefficient it can also be mathematically 'removed' from the Repeatability calculation. It really is surprising how small these Repeatability values can become when thermal drift is removed in this way.

Repeatability is a measure of variability and because of this it also varies in itself. Not only will it vary every time it is measured, but also greater differences will occur over greater periods of time.

It is well-known that measuring Repeatability with a step cycle (where all the individual target readings are taken one after another straightaway) produces 'better' results than a linear cycle which has a longer elapsed time between readings. This is usually a straightforward result of thermal drift as the correlation coefficient will reveal. Extending the time span to minutes, hours, days or even years is a different matter altogether.

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

3 Long Term Effects

It is one thing to measure the Accuracy and Repeatability of a new machine and confirm that the values are within the required test limits. It is quite another to measure it after the machine has been shipped several thousand miles and installed in a customer's plant under quite different operating and ambient conditions. Many customers like to check the machine's accuracy after delivery, either by measuring it themselves or by calling in a specialist calibration company who will carry out laser and ball-bar checks and compare the machine against its original run-off test data. Customers often possess the relatively inexpensive ball-bar and, armed with a PC and a diagnostic software package, they can do quite a lot to evaluate the machine for themselves. In particular, the more enterprising carry out regular checks to monitor the long-term drift in performance as the machine carries out its day to day work in the factory.

Usually, but not always, the customer's shop floor environment will be less conducive to good testing practice than the manufacturer's own premises. And this may influence results in quite a number of ways.

3.1 Angular and Straightness Errors

The accuracy of the machine depends not only on its positioning capability along an axis, but also on the straightness of the axis and the three angular deviations: pitch, roll and yaw. The alignment or squareness of one axis with another is equally as important. It is only with positioning, however, that the concept of Repeatability is used. In BS3800 part 2, which has now been superseded by the ISO, the statistical methods employed for Repeatability of positioning were extended to cover straightness, pitch, roll and yaw. Although this British Standard was an important committee document used by the ISO in preparing their new standard, it was felt that they could not accept that 'Repeatability of Straightness' (for example) was a valid or even meaningful machine parameter, and that such variation that could be measured was not an intrinsic property of the machine, but was probably due to error or uncertainty in measurement. This may be true, as it is difficult to see what could cause variation in these parameters at fixed positions over a short time interval. Over a longer time interval, the situation becomes quite different, as we shall see. Apart from the contentious issue of Repeatability in this different context, however, there did still remain in BS3800/2 a useful measurement procedure for these parameters. With its demise we have now lost this.

This should not preclude us from carrying out such tests, however, as they can often reveal a great deal of useful information about a machine's accuracy. Measurement of yaw is best carried out using a laser system, and a complete axis can be evaluated automatically in a minute or so. With the right software the yaw results can be integrated to generate 'horizontal straightness'. Pitch and roll are probably best evaluated with electronic levels and, again, pitch can be

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

integrated to produce vertical straightness. This method obviates the need to use straightness optics with a laser, which is usually a more lengthy procedure.

3.2 Running-in

When a fully calibrated machine is shipped to a customer and installed in his plant it is not uncommon to discover that its accuracy has changed. Running the machine in will certainly affect its performance. Ballscrews, slide-ways, covers (telescopic or sliding)—and even lip seals on scales—will change their characteristics as they become 'run-in'. Generally, such changes will bring about reduced friction, which in turn will bring about reduced loading. This will always reduce the so-called random effects and thus help to improve Repeatability. The influences of both thermal and hysteresis effects may thus be mitigated by running-in. With most machines the running-in time required for this is quite small, and once this initial phase is over machine accuracy should remain reasonably consistent over a long period.

3.3 Temperature Effects

Operating the machine in a workshop at a temperature different from that prevailing during calibration will certainly make for difficulties. We all know that thermal problems account for much of the inaccuracy encountered nowadays in machines. A machine relying solely on encoder feedback cannot be expected to position to a high degree of accuracy. Ballscrew heating through friction between nut and screw (depending on the amount of work being done by the machine) will result in screw expansion and consequent inaccurate positioning. If the ballscrew is constrained at both ends the effect will be lessened, but not eliminated. When serious accurate positioning is required it really is essential to use scale feedback. Unfortunately, ISO230-2 does not help us to distinguish clearly between performance capabilities of different types of feedback, since it operates under iso-thermal conditions only. The long-awaited ISO230-3[5] thermal test will surely sort the scales out from the encoders! Thermal distortion effects from rotating spindles and consequential structural growth needs to be tackled with active thermal compensation.

Although scales are not so susceptible to machine heating, they still respond to ambient changes. Machines are calibrated so that their scales are accurate at 20degC. At 30degC they will be 84microns per metre longer and positioning will be affected accordingly. The workpiece will have expanded by a further amount; steel for instance will be 117 microns/metre longer. Measuring the workpiece back in the standard environment of 20deg will still result in an error of 33 microns/metre due to differential expansion. If the workshop is always at 30degC the machine can be re-calibrated, but when the shop temperature varies on a short term basis even greater variations in Repeatability will be seen. The only solution would seem to be a highly developed thermal

Transactions on Engineering Sciences vol 23, © 1999 WIT Press, www.witpress.com, ISSN 1743-3533

compensation system that can be fine tuned to match the workpiece material and the ambient effects in factories caused by daily and weekly heating patterns.

3.4 Levelling

An important factor in the installation a new machine is the mundane task of levelling. Surprisingly, this can make a considerable difference to the calibration. Most machines cannot be treated as rigid bodies that can be supported kinematically. They have to be carefully levelled to ensure that all pitch and roll is minimised. A machine that is allowed to sag in the middle through incorrect support will usually have a significantly different positioning accuracy, and its original calibration may no longer be accurate. Often a machine may sag because its foundations have settled due to earth movement. The stress loading in its levelling screws is altered, and pitch and vertical straightness errors develop as a result. Often it is more complex: because these pitch errors are not uniform, twisting of the basic machine structure occurs, which in turn generates the other angular errors including errors of squareness between axes. Squareness may then actually vary at different positions within the work plane.

Because ground settlement is a slow process these errors will also vary gradually with time. Variation with position can be checked with a ball-bar test, which always evaluates *local* squareness, i.e., the squareness within the small circle covered by the ball-bar. Because squareness values can vary with position in the work plane, such values obtained from ball-bar plots may often seem to be curiously at variance with those measured by conventional methods.

4 Conclusions

Despite some earlier misgivings, it would seem that the change to the ISO standard—whether by luck or design—actually does produce sensible figures for Repeatability. Its effectiveness will be further enhanced when the complementary thermal standard is published. And although there is still much work to do in evaluating long-term effects, it is clear that machines have to be installed with great care to maximise the benefits of their laser calibration.

---

[1] ISO 230-2: 1997 Determination of accuracy and repeatability of numerically controlled machine tool axes.

[2] Morris, T.J. The REAL Accuracy of Machine Tools, *Proceedings of 'Lamdamap 97'*, CML Publications, 1997

[3] JIS B 6330, Test code for performance and accuracy of machining centers (vertical type) 1985

[4] ANSI/ASME B5.54, Methods for performance evaluation of CNC machining centers, 1991

[5] ISO230-3 *currently awaiting publication*