Guidance - Accuracy and Measurement Resolution

Three terms are in common use when describing the accuracy of a measurement system.

Resolution

Precision

Accuracy

These terms are often confused and misunderstood, and this note attempts to explain the differences.

Error	The difference between the value of a measurement as indicated by an instrument and the absolute or true value
Sensing Error	An error arising as a consequence of the failure of an instrument to sense the true value of the quantity being measured
Systematic Error	An error to which all the readings made by a given instrument are subject; examples of systematic errors are zero errors, calibration errors and non-linearity
Random Error	Errors of an unpredictable kind. Random errors are due to such causes as friction and backlash in mechanisms
Observer Error	Errors due to the failure of the observer to read the instrument correctly, or to record what he has observed correctly
Repeatability	A measure of the random scatter of successive readings of the same quantity
Sensitivity	The smallest change in the quantity being measured that can be detected by an instrument
Resolution	The smallest difference in instrument reading that it is possible to observe
Average Error	Take a large number of readings of a particular quantity, average these readings to give a mean value and calculate the difference between each reading and the mean. The average of these differences is the average error; roughly half the readings will differ from the mean by more than the average error and roughly half by less than the average error

Error Band

This is the algebraic sum of the total combined errors from hysteresis, linearity, repeatability, regulation, and all environmental parameters expressed as a percentage of the full-scale output.

Theoretical Curve

The theoretical curve is used to determine the magnitude of errors. It is a straight-line plot between 0 volts and maximum volts directly proportional to the 0 percent and 100 percent of the measured parameter. Any deviation from this theoretical straight line is the unit output error.

Static Error Band

The static error band defines the maximum permissible deviation from the theoretical curve and includes the effects of linearity, hysteresis, repeatability, excitation regulation and end points. Any data point should not be greater than the percent of full scale as specified from a corresponding parameter point on the theoretical curve.

Total Error Band (Dynamic Error Band)

The total error band includes all deviations from the theoretical curve due to environment, electrical characteristics, unit performance and any other factors that could contribute to the errors in the measuring system. Any data point should not be greater than the percent of full scale specified from its corresponding parameter point on the theoretical curve.

Resolution

This is the smallest difference in the measurement of a value that can be detected. In modern digital systems resolution is normally defined in terms of the resolution of the analogue to digital converter, which samples the source data. Where the source transducer is digital (e.g. incremental encoders) the resolution is the value represented by one 'bit'.

Commonly used ADC's give the following resolutions :

8 bit 1:256

10 bit1:1024

12 bit1:4096

16 bit1:65536

For example a 10 kN load cell sampled by a 16 bit ADC would have a resolution of 0.305 N (remember that if the load cell operates in tension and compression, the 65536 steps actually cover 20 kN and not 10 kN.

However a stroke transducer covering a range of 100 mm would have a resolution of 1.525 microns since we usually define stroke as the 'total' range.

Digital processing may modify this resolution.

Sometimes the processing may truncate the measurement value (e.g. a 10 bit number may be truncated to 8 bits for display purposes) and the claimed resolution should be modified accordingly.

Conversely it is actually possible to improve the resolution by over-sampling. Consider a situation where a value is sampled with a 16 bit ADC but is subsequently stored and processed as a 32 bit number (range of 4.29×10^9). Simply storing and processing as a 32 bit number does not improve the resolution but repeatedly sampling a value and averaging the result can, if the number range used allows it. For example if we average 64 samples we improve the effective resolution by $\sqrt{64}$ i.e. a factor of 8, which is equivalent to sampling with 19 bits instead of 16 bits. This only works if the source signal has random noise that is greater than 1 bit (1 resolution step) in magnitude. In reality this is usually the case.

An analogue signal can also be defined as having a resolution. This is usually defined as the rms noise level in the measuring bandwidth, as no value can be resolved within the noise band.

Since most digital systems employ some form of filtering (bandwidth reduction) the ADC sampling normally defines the resolution.

Precision

This term although often used has no generally accepted meaning, although many individuals will happily give a definition these definitions will vary!

One consistent definition is in computer science, the term 'number precision' is often used where it defines the number range. In effect it is directly equivalent to the resolution defined above. A number can be defined as having 8 bit precision or 32 bit etc.

Accuracy

The measurement resolution has no direct relationship to measurement accuracy. Just because we can discriminate between two adjacent values does not mean that our measurement is accurate.

Consider a metre rule with graduations at 0.5 mm intervals (a resolution of 0.5 mm) but suppose the rule has been elongated by 10%. We can measure a distance as being, say, 400 mm to within 0.5 mm, however, since the rule is elongated then the true distance is nominally 363.6 i.e. we have an error of more than 36 mm, far greater than the resolution interval.

Accuracy is simply the maximum difference between the measured value and the true value normally expressed as a percentage. Two methods are in common use:

Method 1: Expressed as a percentage of Indicated Value

Method 2: Expressed as a percentage of Full Scale

Method 1 is commonly used in materials testing but it is necessary to define a lower limit e.g. '±1% of indicated value to a minimum of 5% of FS'.

Most transducer manufacturers usually define accuracy using Method 2.

Accuracy errors are a combination of several components :

Non linearity

Offset error/drift

Sensitivity (gain) error/drift

Hysteresis

Non repeatability

These errors may be introduced by the transducer or components of the signal conditioning chain and it is normal in a measuring system to define the total error in the accuracy figure.

Because forces are generally measured relative to a true zero, Method 1 is the more appropriate for definition of force measurement accuracy.

As displacements are generally measured relative to an arbitrary starting point, Method 2 is the more appropriate for definition of displacement measurement accuracy.

Examples of typical accuracies quoted for tribometers are as follows:

Force :	±1% of indicated value to a minimum of 5% of FS
Displacement :	±0.05% of FS

It must be noted that in this case, FS (full scale) represents the calibrated range of the sensor or instrument. It is usually possible to re-calibrate force sensors over a reduced range to enhance the resulting accuracy, for example, calibrating a 10 kN load cell over a 5 kN range, thus reducing FS from 10 kN to 5 kN.

Uncertainty

While it is seldom mentioned in statements of the accuracy of a particular instrument or measurement, the concept of uncertainty is central to any meaningful discussion of accuracy. Uncertainty is a property of a **measurement**, not of an **instrument**:

> **the uncertainty of a measurement is defined as the range within which the true value is likely to lie, at a stated level of probability**

The level of probability, also known as the confidence level, most often used in industry is 95%. If the confidence level is 95% there is a 19 to 1 chance that a single measurement differs from the true value by less than the uncertainty and one chance in 20 that it lies outside these limits.

If we make a very large number of measurements of the same quantity and plot the number of measurements lying within successive intervals we will obtain a distribution curve. The corresponding theoretical curve is a normal or Gaussian distribution. This curve is derived from first principles on the assumption that the value of any "event" or measurement is the result of a large number of independent causes (random sources of error).

The normal distribution has a number of properties:

> the **mean value** is simply the average value of all the measurements
>
> the **deviation** of any given measurement is the difference between that measurement and the mean value
>
> the **standard deviation** σ (sigma) is equal to the square root of the sum of the squares of all the individual deviations

The standard deviation characterises the degree of "scatter" in the measurements and has a number of important properties. In particular the 95% confidence level corresponds to a value $\sigma = 1.96$. 95% of the measurements will lie within these limits and the remaining 5% in the "tails" at each end of the distribution.

In many cases the "accuracy" of an instrument as quoted merely describes the average value of the deviation i.e. if a large number of measurements are made about half will differ from the true or mean value by more than this amount and about half by less. Mean deviation = 0.8 σ approximately.

However this treatment only deals with the random errors: the systematic errors still remain. To give a simple example consider the usual procedure for checking the calibration of a force transducer. A calibration arm, length 1.00 m carries a knife-edge assembly to which a dead weight of 10.00 kg is applied. The load is applied and removed 20 times and the amplifier output recorded. This is found to range from 4.935 V to 4.982 V with a mean value 4.9602V.

The 95% confidence limit for a single force reading may be derived from the 20 amplifier output readings and, for the limiting values assumed, would probably be about ± 0.024V, or ± 0.48%, an acceptable value.

There are however four possible sources of systematic error:

the local value of g may not be exactly 9.81 m/s

the mass of the dead weight may not be exactly 10 kg

the length of the calibration arm may not be exactly 1.00m

the voltmeter used may have its own error

In fact, none of these conditions can ever be fulfilled with absolute exactness.

Repeatability and Reproducibility

Many test standards (ASTM in particular) will include values for repeatability and reproducibility based on inter-laboratory tests. These are typically of the form:

> **Repeatability** — The difference between successive results obtained by the same operator with the same apparatus under constant operating condition on identical test material would, in the long run, in the normal and correct operation of the test method exceed the following values only in one case in twenty.
>
> **Reproducibility** — The difference between two single and independent results obtained by different operators working in different laboratories on identical test materials.

The repeatability value is in fact the 95% confidence level for the particular experiment.

In addition to repeatability and reproducibility, the procedure will also include a bias statement, typically of the following form:

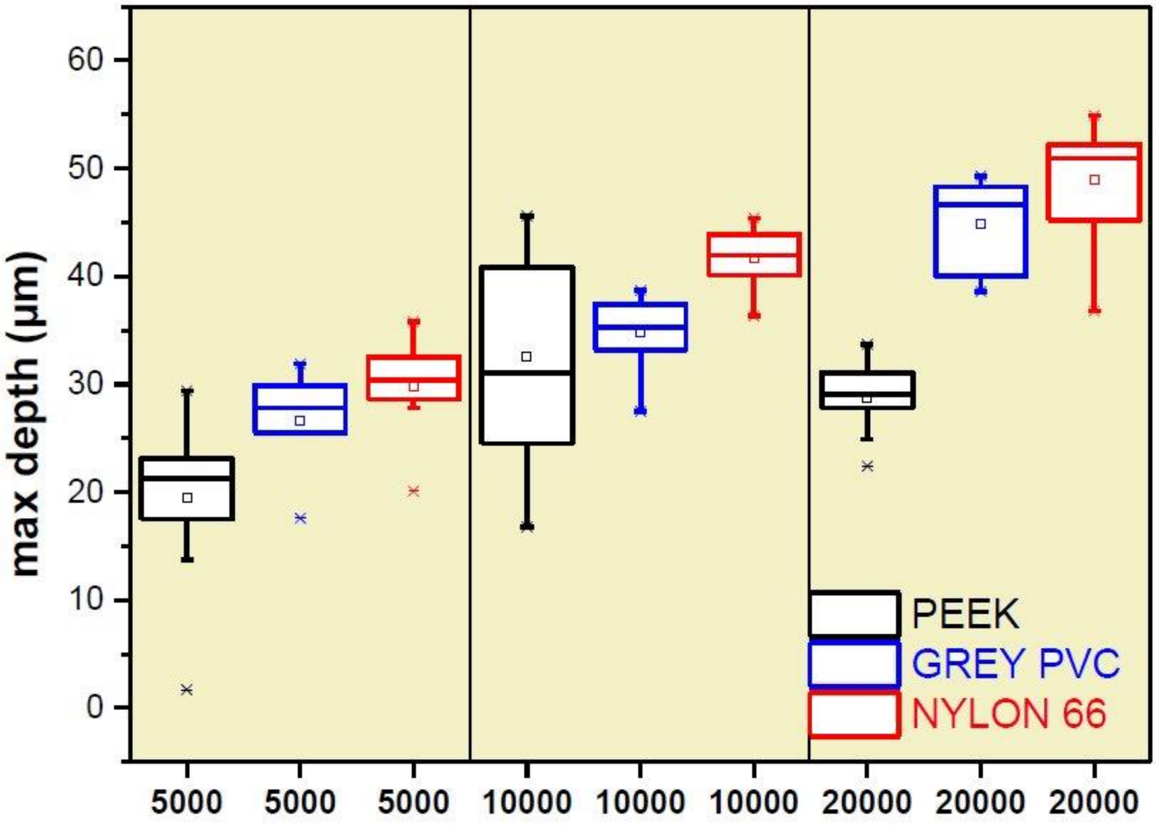
> **Bias** — The evaluation of friction and wear properties by this test method has no bias because coefficients of friction and wear can be defined only in terms of the test method.

In essence, this statement implies that the measurements made with the given method produces results that do not correlate with results generated using other test methods or indeed the performance of the materials tested in other applications.

Testing for Outliers

An outlier is a measurement that appears to lie an abnormal distance from other values, in a series of repeat measurements. Various techniques are available for identifying outliers, of which the most commonly used in tribological experiments, is the box plot.

The box plot is a graphical display indicating the behaviour of data at the middle and ends of the distribution, using the median and lower (Q1) and upper quartile (Q3) of the distribution. The difference (Q3 - Q1) is defined as the interquartile range (IQ).



Example Box Plot

Wear (ten repeats) after 5,000, 10,000 and 20,000 revolutions

Box plot “fences” are constructed by drawing a box between the upper and lower quartiles with a solid line drawn across the box to locate the median. The following “fences” are used to identify extreme values in the tails of the distribution:

lower inner fence: Q1 - 1.5*IQ

upper inner fence: Q3 + 1.5*IQ

lower outer fence: Q1 - 3*IQ

upper outer fence: Q3 + 3*IQ

A point beyond an inner fence on either side is considered a mild outlier. A point beyond an outer fence is considered an extreme outlier. It will be apparent that if a decision is made to discard extreme outliers, there will be a corresponding improvement in the re-calculated confidence level.

Standard Control and Data Acquisition Systems

For standard control applications using Phoenix Tribology’s standard USLIM Serial Link Interface Module, the ADC resolution for both Control and Data Acquisition is 12 bit, with a maximum control output and sampling frequency of 12 bits.

For applications using Direct Digital Control of a motor via a.c. vector drive or a.c. servo motor amplifier, the control resolution will be that specified by the manufacturer of the drive. This is typically 16 bit.

For high-speed data acquisition applications using Phoenix Tribology’s standard HSD data acquisition card, the resolution is 16 bit.

Calibration of Force and Torque Transducers

The largest systematic error with most tribological tests is associated with the use of statically calibrated force or torque measuring transducers and their associated specimen test assemblies.

THERE IS NO NATIONAL OR INTERNATIONAL STANDARD SPECIFYING THE DYNAMIC CALIBRATION OF FORCE OR TORQUE TRANSDUCERS

Until this issue is resolved, experimenters and suppliers of equipment will continue to use STATICALLY calibrated force and torque transducers to measure DYNAMIC forces. The best that can be hoped for is that experimenters are aware of this problem and hence aware of the potential limitations with regard to comparison of results generated on different test machines operating with nominally the same test geometry. The following should be considered:

Resonant Frequencies

The higher the stiffness and the lower the mass of system components, the higher the natural frequency and the frequency response of associated measuring systems. Hence, stiff, light weight designs, with low inertia will be required for all but very low speed applications, in order to avoid resonant frequencies problems.

With high moving mass systems such as servo hydraulic actuators and large electro-magnetic vibrators, there is a requirement for the overall test machine to have a high mass in order to minimize machine vibrations. Force measuring systems with associated tooling act as very effective accelerometers, with the risk that the true friction signal will be swamped by parasitic vibrations.

Frequency Response

The information content available in the signal channels of a dynamic testing machine is directly related to the signal bandwidth. The fundamental limitation in most measuring systems is the bandwidth of the transducer itself.

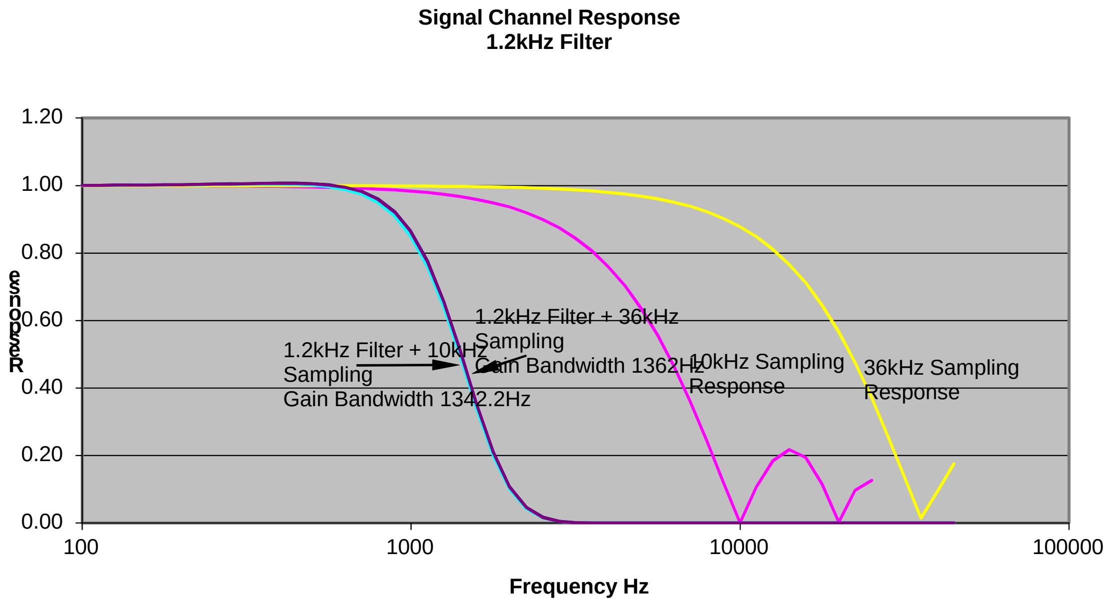
In most testing machines it is the load cell that will be the limiting factor in the form of its mechanical resonance. Most low-profile strain gauge load cells will have a natural resonant frequency in the range 4000 Hz to 5000 Hz with an internal effective mass typically in the region of 0.5 kg. When typical external specimen adapters are added we can expect this resonance to fall to about 3 to 3.5 kHz. It is generally accepted that, in order to keep measuring errors low, a load cell should not be used at frequencies above about 0.3 of its resonance, therefore input filtering is imposed to limit the signal bandwidth accordingly.

A filter should normally be applied to force measuring channels in order to eliminate higher frequency signal noise and aliasing. The same characteristic filter should be used on all channels, especially in high frequency systems, to ensure that measured information can be directly correlated and is not subject to differing time delays.

When sampling the signals processed by such a filter it is important to sample at high enough a rate to preserve the information in original signal. The Nyquist Sampling Theory indicates that the minimum acceptable sampling rate is twice the maximum frequency of interest. The measured amplitudes of signals at half the sampling rate are attenuated to 64% of their true value and it is good practice to sample at higher rates wherever possible. With a system usually sampled at 10 kHz, a 1 kHz signal, which would have suffered 36% attenuation with 2 kHz sampling, only suffers 2% attenuation. The graph shows the differences between the combined effects of a 1.2 kHz input filter and the attenuation effects of 10 kHz sampling and 36 kHz sampling. The differences are barely noticeable.

Estimating the gain bandwidth product (an estimate of the information carrying ability of the system) shows that the 36 kHz sampled system is only 1% better than the 10 kHz system. In other words, sampling faster than necessary may produce more data but no more information! We could obtain the same amount of data by using a lower sampling rate and simply interpolating between the samples.

To conclude, the Sampling rate of the system should be well matched to its signal bandwidth in order to preserve information content. Furthermore, the signal bandwidth should be well matched to typical measuring transducers.



Temperature Measurement

Phoenix Tribology Ltd supplies machines fitted with standard Class 2 thermocouples. It is worth noting that Class 2 thermocouples, J, K and N type, are accurate to only +/- 2.5 °C, Type T to +/- 1.0 °C and type R to +/- 1.5 °C. We use type K in most application, but then the standard allows a temperature range of +/- 5 °C!

To go beyond the calibration accuracy of the standard thermocouple, we have to calibrate each thermocouple against a temperature reference. In theory, the thermocouple should be re-calibrated every time it is disconnected/reconnected from the thermocouple plug, as the plug connections act as junctions in their own right. Thermocouples are much less accurate than people like to think or claim, especially if they have not been calibrated.

It is worth noting that most components and test fluids within dynamic test machines are subject to significant temperature gradients. It follows that positioning and re-positioning after removal of thermocouples and other temperature sensors is critical if random errors are to be avoided.

Conclusion

It should be apparent that, whereas it is possible to define the resolution, precision and accuracy of a given sensor and measurement system, it is not possible to define the resolution, precision and accuracy of some property, measured with that sensor, in some experiment. Think, for example, of a simple hardness test: we may have a very accurate force sensor, but if our indenter is worn or damaged, we will get an inaccurate answer.

The issue with tribological measurements is that repeatability and reproducibility are a function of the materials being tested and the experiments that are performed.

Friction, as measured using nominally identical samples and experimental procedures, is not necessarily, in itself, a very repeatable property; in other words, there is significant natural variability in the friction between mating surfaces.

Referring to the example box plot, it will be apparent that not only do some materials appear more “repeatable” than others, but that the “repeatability” varies depending on the number of cycles run.

When it comes to experiments, some contact geometries and some test procedures produce more repeatable friction and wear data than others, hence the different repeatability and reproducibility statements in different ASTM standards. Friction and wear are system responses, so are not only dependent on the materials being tested, but on the test system and experiment.

It is important not to confuse the precision, resolution and accuracy of a measuring sensor with the precision, resolution, accuracy, repeatability and reproducibility of some physical property measured using that sensor. The only way to establish the “accuracy” of a measurement is to perform sufficient repeat tests to establish an appropriate confidence level.