Practical Considerations for the Use of Double-Flank Testing for the Manufacturing Control of Gearing - Part II

Ernie Reiter and Fred Eberle

Part I of this paper, which appeared in the January/February issue of Gear Technology, described the theory behind double-flank composite inspection. It detailed the apparatus used, the various measurements that can be achieved using it, the calculations involved and their interpretation. The concluding Part II presents a discussion of the practical application of double-flank composite inspection — especially for large-volume operations. It also addresses statistical techniques that can be used in conjunction with double-flank composite inspection, as well as an in-depth analysis of gage R&R for this technique.

Statistical Techniques for Double-Flank Inspection

Statistical process metrics are very important to production gears made in large quantities. Unlike small sample runs where nearly every part would be inspected, it would be impractical to do so for high-volume generated or molded gears. As a result, only a few samples are measured at a logical predetermined frequency to monitor the process in time. The primary goals of statistical techniques will determine if the process is — or continues to be — stable without anomalies and to ensure that the manufacturing process is capable. If these are satisfied, then there is confidence that the part quality is also good.

Determining the process capability of total composite variation (error/TCE), tooth-to-tooth variation (error/TTE), and tight mesh center distance (TMCD) or test radius (TR), using double-flank composite inspection (DFCI) is a practical and economic way to determine if a gear manufacturing process is stable and acceptable. This is true for all types of gears, but particularly true for high-volume-production gears such as those molded by plastic injection molding or powder metal processes.

Using DFCI data for statistical analysis can greatly aid in optimizing gear tools during the debugging phase of the initial manufacturing process. It is also an effective means of maintaining ongoing and real-time process control. This section provides an illustrative case study of the use of these statistical techniques during the development stage of an insert-molded plastic gear that is molded on to a steel shaft. The techniques presented are especially suited for initial and ongoing statistical analysis in a high-volume manufacturing environment. In addition, this paper will also discuss two ways of executing a study to validate the accuracy of the measurement system itself.

A Case Study of Statistical Techniques during Gear Development

The gear materials used, the type of injection gating, the uniformity of wall sections, and thermal mold flow behavior all have a significant impact on the resultant geometry and overall quality characteristics of a plastic-injection-molded gear. Experienced gear molders understand these issues and incorporate counter measures to minimize the impact of known causes of variability. However, after the gear molding tool has its initial pilot run and a higher level of gear quality is still required, the inserts and other critical aspects of the mold may need fine-tuning in order to bring select qualities into conformance with design intent. In every type of gear manufacturing operation, specific tooling and process adjustments can be effectively made only from a stable process. When there is a lack of stability, making tooling adjustments is analogous to trying to hit a moving target — rarely successful in tuning gear tools.

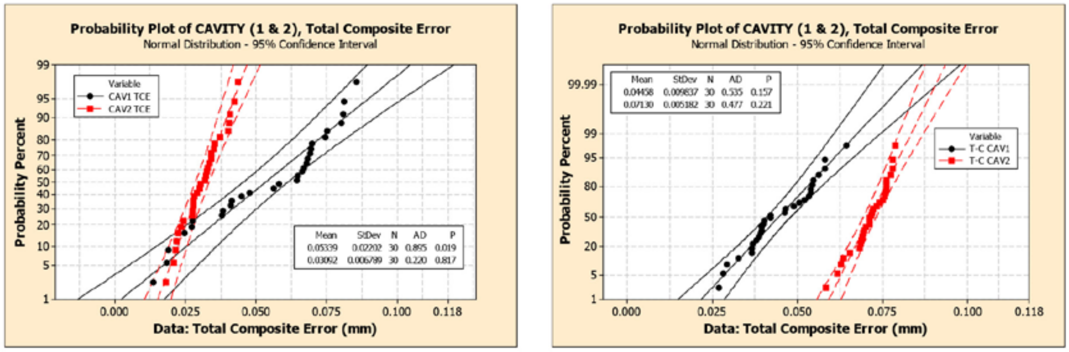
In a total composite error case study of a new tool for an insert-molded plastic gear (Fig. 8), all of the samples measured from the two cavity molds were comfortably within the design tolerance of 0.118 mm. Without looking at the potential process capability, one may have been satisfied with the results, and the tooling could have been approved for production. However, it was discovered using a probability plot (Fig. 9a) that the probability statistic (i.e., the "p-value") at a 95 percent confidence interval for Cavity 1 is 0.019, while for Cavity 2 the value is 0.817.

Figure 8 Over-molded plastic gear on a metal shaft for the case study (all images courtesy of Hi-Lex Controls Inc.).

Note that in creating this chart the null hypothesis of the test is as follows: variation in the sample is normal and hence random in nature; a high probability statistic result means that the hypothesis is more likely correct, while a low result means that the hypothesis is less likely to be so.

The high p-value significance test of Cavity 2 (i.e., $0.817 > \text{alpha}$ value of 0.05 at a 95 percent confidence interval) indicates that the data in Cavity 2 is normally distributed and has random process variation. The low p-value of Cavity 1 shows just the opposite; i.e., that there must be an assignable cause for the variation. We can also see this visually (Fig. 9a) from the variation and scatter within the hyperbolic confidence interval lines of Cavity 1. Note that data approximating normality will have a tendency towards forming a straight line in a probability chart.

In this case study the molder was alerted that variation coming from Cavity 1 was unstable, with an undetermined but assignable cause. Further investigation revealed that the metal shaft being overmolded onto the gear was hanging up slightly in the mold and causing unpredictable distortions of the gear in Cavity 1. After correction, the revised probability statistic was re-evaluated and the acceptable result is shown (Fig. 9b). Note that after the fix, the resulting overall Cavity 1 TCE standard deviation was lowered to 0.00518 mm — a reduction by over 55 percent from the original trial. The resulting reduction in variability — even with this small number of samples — gives statistical confidence that the tool and process are greatly improved and should produce in-specification parts for the long-term. Without this statistical analysis, at some future point during production a vexing quality problem may have surfaced requiring 100 percent inspection of this characteristic. Indeed, the true root cause may never have been found if a probability plot had not been generated.



a) before mold correction

b) after mold correction

Figure 9 Probability plot for the case study plastic gear:

With the special cause of molding variation eliminated, a capability analysis can now be done to determine the strength of the manufacturing process. However, it must be noted that until the total composite error data has a reasonable conformance or approximation of normality, or until a much larger sample of data is taken, a capability study would not be appropriate; it would lack sufficient statistical confidence and accuracy. The revised probability plot (Fig. 9b) shows that the p-value for Cavity 1, with p-value > 0.05, meets the criteria of approximating normality and indicates that the process is stable. If normality exists in small datasets of 30 to 60 samples, capability analysis can be shown to be accurate within its confidence interval. But if the data does not approximate normality (i.e., p-value < 0.05), more data may overcome this condition.

Therefore, now that the process is under control, a snapshot of process capability can be taken. A Cpk (capability performance) analysis (Fig. 10) can be done. In performing this analysis for total composite error, a lower bound is established at the lower specification limit of 0.0, while the upper specification limit is set at 0.118 mm. The lower-bound method is used instead of a lower specification limit because total composite error (as well as tooth-to-tooth error) is always a one-sided, maximum value. A value of less than 0.0 is impossible to achieve. Hence it is only appropriate to calculate capability for both Cpk and Ppk (performance capability) against the upper specification limit.

Furthermore, the mean line on the chart in Figure 10 is only a visual cue for process drift. By showing an optional mean line at the center of the total composite error specification, the analyst can visually see an orientation of shift or drift in future process measurements.

The capability snapshot analysis shows a Cpk of 2.97 with a 0.118 mm upper-specification limit for the gear's total composite error. A minimum Cpk of 2.0 is usually desired for initial and unknown, long-term process data in order to theoretically allow for a 1.5 Sigma drift resulting in a Ppk of 1.67. Based on that assumption, and the data presented here, it is reasonable at some future point to consider reducing the specified tolerance to 0.095 mm, resulting in a Cpk of 2.0 if a stable process and more data bear that out.

NOTE: The Cpk is based only on one snapshot of the process of 30 data points, i.e. — one sub-group. Thus this data represents a preliminary screening test of the process at one point in time. Calculation of the potential (within) capability is a special case for a sub-group = 1, and is done

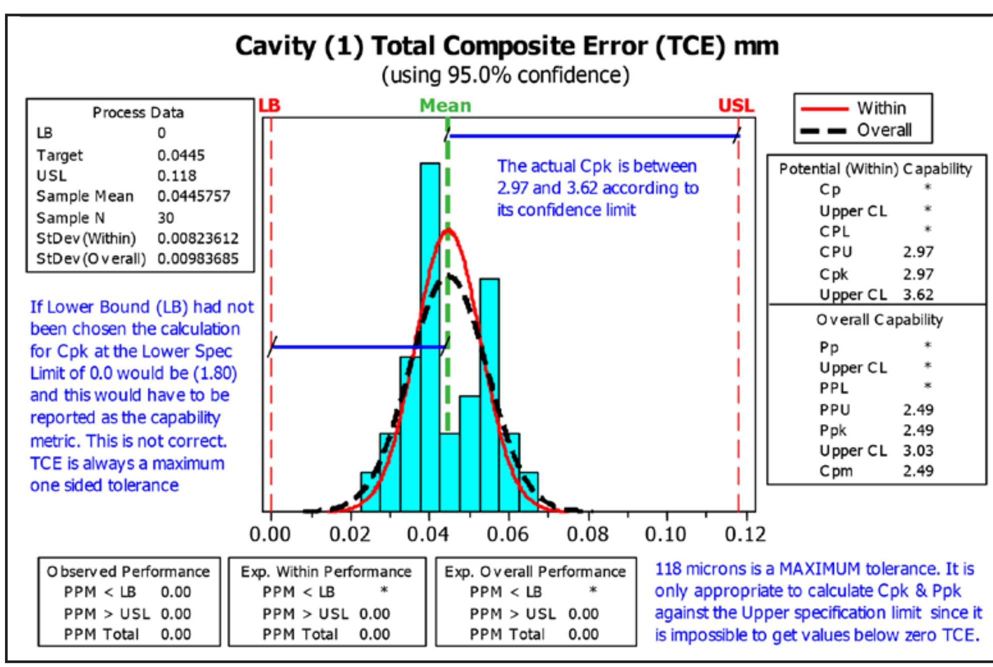


Figure 10 Process capability plot for the case study plastic gear.

with an estimate of standard deviation made with a moving range equation (Ref. 2). Also, Ppk — the long-term estimate of process capability — in this example is determined by inferring a percentage of process drift that is purely theoretical; it is not based on any hard data from this gear. Thus, it is important to understand that in order to truly measure your process, more data must be taken over time — e.g., set-ups, shifts, lots or whatever the case may be — in order to properly assess the condition of the part.

Based on this first production run data, it would be useful to evaluate five to ten sub-groups per-lot of future data. When there are three-to-five production lots established, the re-calculated process capability will lend much more confidence and understanding of the long-term capability of the tool, process, and true quality level of the manufactured gear. A probability distribution plot of total composite error (TCE) (Fig. 11) can be very helpful in predicting the range of the tolerance that the established process will use. It provides more insight into how reliable and robust the actual process is — and is expected to be — in the future. In this case it is predicted that only four percent of TCE measurements in the future will be greater than 0.059 mm or 50 percent of the TCE tolerance of 0.118 mm.

Please note: the 0.059 value is arbitrarily chosen.

With a process in statistical control, total composite and tooth-to-tooth variation will conform to a normal distribution. However, if normality is not present, a look at the histogram will reveal clues as to what shape of parametric or non-parametric distribution the data conforms to. If necessary, other standard statistical tools or transformation techniques are required to normalize the data or create a non-parametric predictive model of the process. It can be shown that as more measurement data is accumulated, the normality requirement becomes less and less important in determining process capability for the metrics described in this paper.

A question that often arises is whether capability requirements with double-flank inspection are necessary if parts are in specification. For high-volume products the customer must be certain that the gear manufacturer is capable of supplying parts within specification, shipment after shipment, and that the manufacturer's process is robust. Most gear purchasers can only verify small quantities of gears using their receiving inspection procedures, yet the consequences of quality issues may be technically and economically catastrophic. In one case study, a high-volume gear sold for $1 and was assembled into an automotive device that sold for $120. The sub-assembly replacement cost of a defective gear in the manufacturing plant was $26, but escalated to over $200 in the field. The economic damage to a customer can be exponentially greater due to the value-added assembly. Process control and capability measurement are crucial in long-term product quality and customer success. Responsibility for this success falls not only on the gear manufacturer, but also the gear designer, who has to ensure that product specifications are reasonable for the intended manufacturing process deployed.

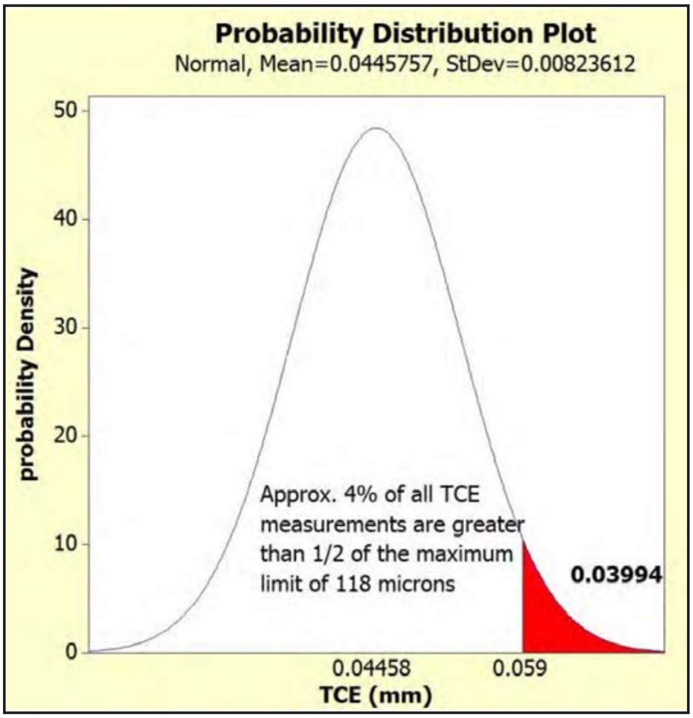


Figure 11 Probability distribution plot.

Statistics for ongoing, double-flank composite inspection. The following documentation applies to molding plastic gears; however, many of the issues described are transferrable to any gear manufacturing processes.

When large quantities of gears need to be produced, it is essential to know where and when quality characteristics related to gear performance occur, and how to use this information to improve your process. Certain statistical analyses are effective at pointing to causes of variation.

For example, assignable causes from DFCI data could indicate:

- A shift in tooth size (evaluate with tight mesh center distance)
- A change in eccentricity or tooth profile conformance (evaluate with total composite error)
- A change in nicks, burrs, handling damage or tooth form (evaluate with tooth-to-tooth composite error)

Applying ongoing statistics to total composite error. Probability curves and capability metrics are useful tools to analyze total composite error. The probability curve opens a window of understanding on the type of variation present in double-flank roll data. Similarly, it is useful for determining if the data is stable and suitable for making potential capability models and conducting certain hypothesis tests. If the probability analysis is acceptable, then a capability study can be prepared specifically to:

- Assess the potential strength (robustness) of a process at a specific point or points in time
- Predict the future potential of a process to create a values within design limits using meaningful metrics
- Identify improvement opportunities in the tooling or process by reducing or possibly eliminating sources of variability (Ref. 2).

Ongoing, TCE capability evaluation provides meaningful insight into the robustness of the products being made and confidence in the ability to consistently make parts to design intent.

Applying Ongoing Statistics to Tooth-to-Tooth Composite Error

Tooth-to-tooth error is best scrutinized from real-time tracking on an Xbar-R (average and range) chart for data in sub-groups $<=8$, or an Xbar-S (average and standard deviation) chart for sub-groups >9. For these larger sub-sets, the range in an Xbar-S chart is a much better statistic to estimate distributions of the sub-groups.

An Xbar-S chart is usually the most appropriate choice to detect changes in tooth-to-tooth error for molded gears with a reasonable sensitivity. A real-time, in-process TTE control chart is highly recommended because changes in tooth-to-tooth error are generally the first indicator of a change in the manufacturing process.

The change in tooth-to-tooth error, particularly in injection molding — but also in powder metal gearing, fine-blanking and wrought gear cutting and grinding — will signal potential handling damage, such as burrs (or flash); nicks; part ejection issues; dirty or worn tooling; and other surface irregularities. When these special-cause events start to occur, the TCE will not have nearly as much sensitivity as tooth-to-tooth error. TTE will have more sensitivity in flagging these events over TCE due to its smaller tolerance.

The goal in control charting is to capture any degradation in gear quality and resolve it before the manufacturing process trends out of control. In addition to in-process verification, control charts can be a particularly important tool to validate set-up and process conditions just prior to approving large production runs. The data does not need to be normal, but if the process is not stable, any chaotic behavior will be graphically evident and the control limits may not be valid.

When using control charting, tooth-to-tooth data, it is important to:

- Take at least 100 data points to ensure that the control limits are precise, or else consider the results preliminary.
- Evaluate the data in time order to identify trends as they occur; control charting after production precludes any possibility that a process correction could be made in time to save or improve the production run.
- Select an appropriate number of measurements for individual sub-groups using the following equation:

$$N = \left[\frac{(N_{\alpha/2} + N_{\beta})}{D} \right]^2 \tag{24a}$$

where:

N is the number of measurements needed for an individual subgroup

$N_{\alpha/2}$ is the number of standard deviations above zero in a normal distribution — ±3 Sigma is commonly used.

N_{β} is based on a percentage probability that the analysis will detect a shift in standard deviation between sub-groups (note that for an 80 percent probability, $N_{\beta} = 0.84$)

σ is the historical standard deviation established for the process; this parameter must be historical to a particular tool, process or cavity of interest; the value is important and should be appropriately established.

D is the sensitivity parameter; i.e., the amount of change in tooth-to-tooth error that is to be detected.

For example, given a historical tooth-to-tooth error standard deviation of 0.0045 mm, and an 80 percent probability that a shift of standard deviation can be detected within 0.0055 mm, determine an appropriate number of measurements needed for individual control chart sub-groups using ±3 Sigma. Using equation 24 —

$$N = \left[\frac{(3 + 0.84)\,0.0045}{0.0055} \right]^2 = 9.87 \tag{24b}$$

Therefore a minimum of 10 measurements-per-sub-group should be used, and an Xbar-S chart (Fig. 12) is most appropriate. Over the course of 100 measurements (Fig. 12) the control chart shows that the process is consistent and predictable, with only random variation present. A look at the sub-group means in the top chart gives a good reading of where the process is and the average sub-group, tooth-to-tooth error spread between sub-groups. In the lower chart, even though there are some numerical differences between sub-group standard deviations, all standard deviations fall within the calculated control limits.

NOTE: Never consider the calculated control limits as relative to component design specifications — they are not related. The upper- and lower-control limits are based only on the measured sub-group data. The expected or predicted variation in the process is calculated as 3 Sigma above and below the average sub-group standard deviation line. The control limits are used to determine if any sub-group behaves differently than expected. The control chart does not tell us if the process is capable within the specified tolerances; it only tells us if the process is trending or stable within the calculated control limits. Capability relative to its tolerance specification is a separate statistical check, as previously discussed.

Applying ongoing statistics to tight mesh center distance or test radius. Tight mesh center distance (or test radius for the purpose of this section) are often regarded to be significant product characteristics on gear drawings. For example, in injection molding of plastic gears the tight mesh center distance data measured during double-flank inspection is affected by cavity pack pressure. Higher pack pressures generally increase tooth size and reduce total composite error. Lower cavity pressures increase total composite error and reduce gear tooth size.

Even though the tight mesh center distance results include the effects of total composite error and tooth thickness, it is some-

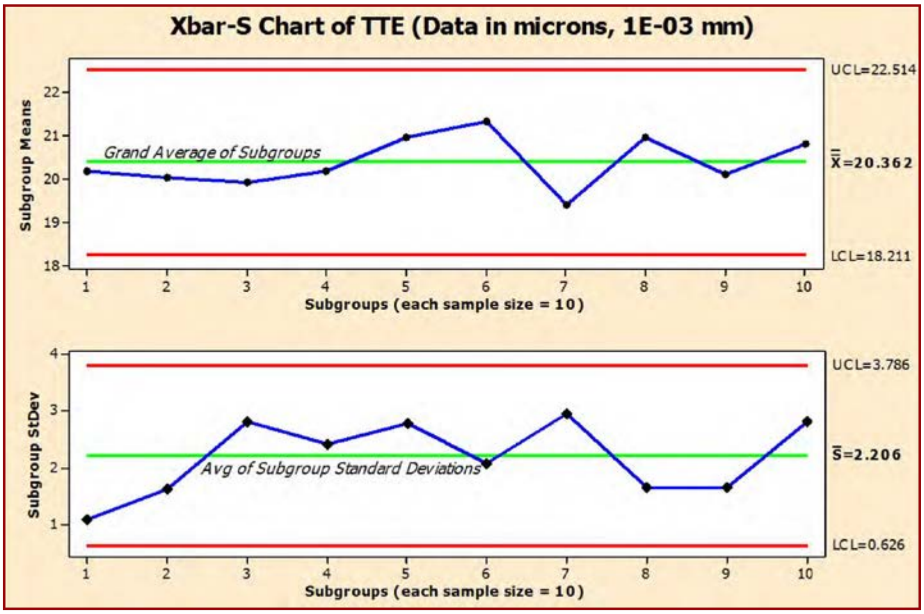


Figure 12 Typical control chart of in-process tooth-to-tooth error data.

times problematic to deal with this parameter in a statistical manner.

If a capability requirement is specified on tight mesh center distance, the procedure for applying capability is extremely important. This is because an individual gear does not have a single parameter of tight mesh center distance. As an example, the gear measured in Figure 3 has a maximum and minimum tight mesh center distance reported as 23.115 and 23.093 mm, respectively, resulting in a mean value of 23.104 mm. For the individual part to be in tolerance, every rotational position of the gear must be in specification when rolled against its master. Therefore if only a single value for tight mesh center distance is assigned to an individual part (mean value, for example), there are no functional specification limits to calculate its capability.

In order to calculate Cpk or Ppk on tight mesh center distance, two separate capability studies must be performed (Figs. 13–14). The first study uses the minimum TMCD data in comparison to the minimum specification value. By setting an upper-bound on the maximum TMCD specification, the capability metric is guaranteed to be taken against the lower specification value only. The second half of the study uses the maximum TMCD data against the maximum specification value. This time the minimum TMCD specification value is selected as a lower-bound to guarantee that the capability metric is only taken against the maximum specification value. The actual calculated Cpk or Ppk will be the lesser of the two reported values; in (Figs. 13 and 14) the actual TMCD capability Cpk = 0.57.

NOTE: Using the upper- and lower-bound method reports a Cpk or Ppk as being towards the desired specification only. Had the upper-bound method not been used (Fig. 13), the minimum tight mesh center distance, Cpk, would have been incorrectly reported towards the upper specification, since the data is closer to the upper specification limit. In (Fig. 14) the use of the lower-bound made no difference to the final result since the actual data is in fact closer to the upper specification.

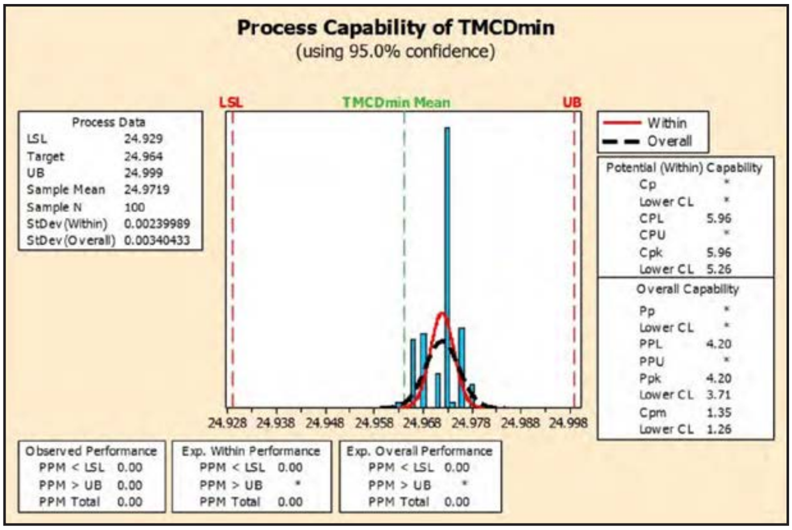


Figure 13 **Minimum tight mesh center distance capability study.**

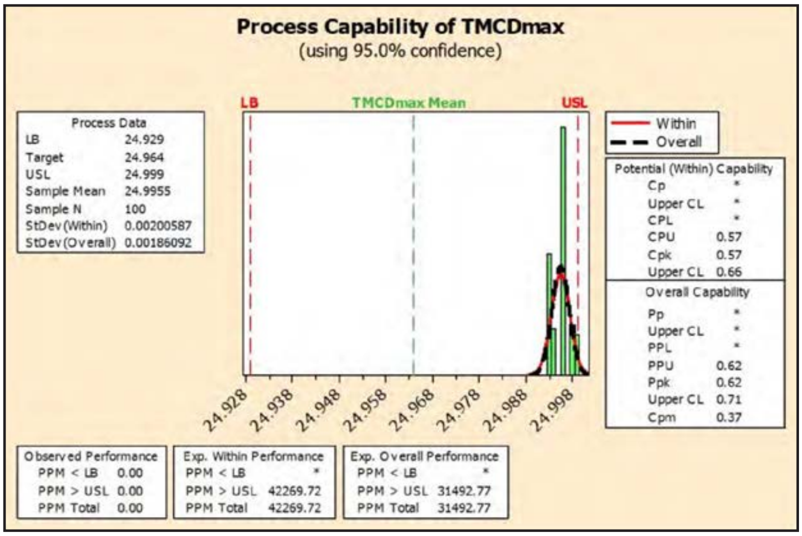


Figure 14 **Maximum tight mesh center distance capability study.**

From the figures it can be seen that although the minimum TMCD is capable, the maximum TMCD is not; hence, a process or tooling shift is needed. A subsequent reduction in the tooth thickness (in this case by 13 microns smaller) resulted in shifting both the maximum and minimum tight mesh center distance results lower by 18 microns towards the lower specification limit. This centering shift results in an improved overall Cpk of 3.46.

In this case a satisfactory result was obtained with only a shift in the tooth thickness. In some instances it is possible that, even using the best tooling and processing analysis, a further improvement in the capability index may not be possible. In such cases it may be necessary to question the suitability of the tight mesh center distance tolerance and the parameters that make up that tolerance (i.e., tooth thickness variation and total composite variation). If the total composite error is in statistical control, one would have to conclude that the tooth thickness variation itself is the issue that must be reviewed.

Measurement System Analysis of Double-Flank Composite Inspection

The most common method of measurement system analysis used in manufacturing at this time is gage repeatability and reproducibility (gage R&R). There are inherent issues in double-flank testing due to the dynamic function of the gage, which typically results in higher gage R&R results than what is typical for other static-type measurements. As such, gage R&R may not be the most suitable approach for measurement system analysis when it comes to double-flank composite inspection practices. Instead, uncertainty analysis may be a more fitting approach for measurement system analysis, as will be further explained.

Gage R&R and double-flank composite inspection. By definition, gage R&R by ANOVA method (*Editor's Note: analysis of variance — ANOVA — is a collection of statistical models used to analyze the differences between group means and their associated procedures, such as "variation" among and between groups*) is the amount of measurement variation introduced by a measurement system that consists of the measuring instrument itself and the individuals using the instrument. A gage R&R study quantifies three things (Ref. 2):

1. Repeatability: variation from the measurement instrument

2. Reproducibility: variation from the individuals using the instrument
3. Overall gage R&R: the combined effect of 1 and 2

The gage R&R study determines how much of the observed process is due to measurement system variation. It breaks down the overall variation into its part-to-part repeatability and reproducibility components. The total gage R&R is the sum of all the study variation minus the actual dimensional variation between parts. The goal is that 90 percent or more of measured variation be due to the actual dimensional differences in the study parts; it is desirable that only 10 percent of the variation be attributed to the repeatability of the gage and appraisers.

The two-way ANOVA table lists the following sources of variability (Fig. 15):

- Part: Represents the variability in measurements across different parts. The goal is that all the variation be identified as the actual size differences between parts.
- Operator: Represents the variability in measurements between inspectors.
- Operator * Part: Represents any potential interactions between these two main effects.

The overall gage R&R is normally expressed as a percentage of the specified tolerance for the attribute being studied. A value of 20 percent or less gage R&R is considered acceptable in most cases.

Gage R&R case study results. The following is a real example gage R&R on total composite error that was carefully performed by three highly trained appraisers on five identical, unfilled acetal plastic gears tested against a high-quality master gear. The equipment, electronic controls and software used are considered to be some of the very best the industry offers. A well-made, unfilled acetal gear was chosen because acetal is one of the most common and accurately molded plastic gear materials used in industry.

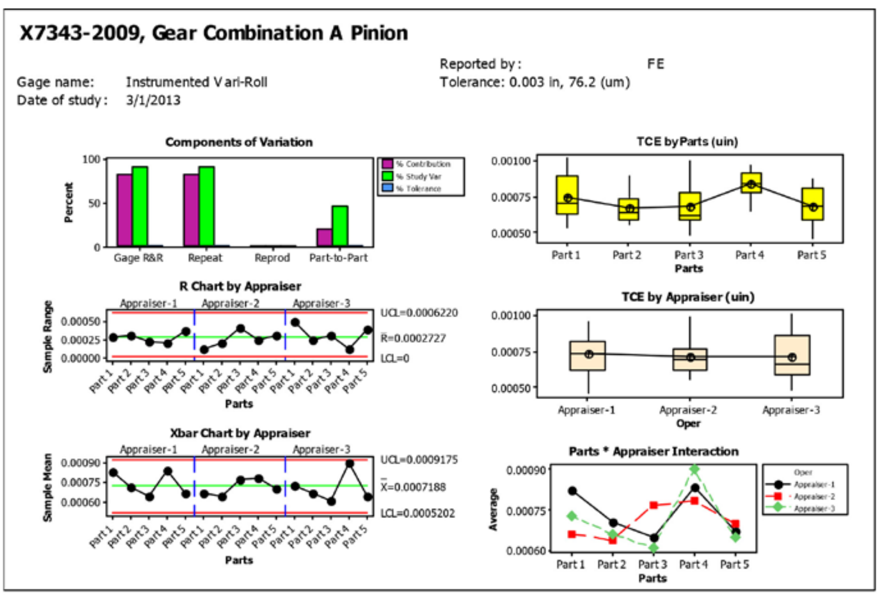
The analysis (Fig. 15) shows that the total gage R&R fails at the 80.3 percent level, and that part measurement repeatability is the primary contributor with over 80 percent of the total variation contribution, as compared to under 20 percent coming from part-to-part variation.

As stated previously, according to the ANOVA method, repeatability is related to the measuring instrument. This result, however, seems illogical given the quality of the instrument employed and the extremes which the study went to in order to avoid variation in the instrument and measurement process. These methods included motorized rotation of the gearing, computerized data collection and high-quality part holding methods to ensure centering of the part on the measurement mandrel.

A very revealing value shown in the ANOVA table (Fig. 15) is the number of distinct categories value of 1. For a measuring system analysis with five parts, a value of five distinct categories is typical. A value of 1 would mean that the measurement system is of no value in evaluating the process, since one part cannot be statistically distinguished from another part.

Furthermore, in the Xbar chart (appraiser section, Fig. 15) it is shown that there is variation within the same part from appraiser to appraiser, while the "Part * Appraiser Interaction" curve shows in a different way how far apart the appraiser variation is from each other. Since the p-value for the potential Part * Operator interaction is 0.413, and hence greater than 0.25, we conclude that there is an interaction between part and appraiser, and that this statistic will be included in the error calculation and not be dropped from the gage R&R model (Ref. 3). This interaction, as reported by the ANOVA method, again seems illogical given the extreme control exercised in taking the measurements and the training and experience of the operators involved.

Further consideration of this issue was given by virtue of observation over many different types of gears including plastic, powder metal, fine blanked, and wrought cut gearing. The question of what effect can arise from minor surface imperfections in a gage R&R was considered. Surface-related issues that could affect measurements like dirt, cutting fluids, burrs, nicks, parting line flash and wear may have an impact on results from one



Gage R&R Study - ANOVA Method for Total Composite Error

Gage name: Vari-Roll
Date of study: 3/1/2013 Reported by: FE
Tolerance: 0.003 inch

Two-Way ANOVA Table With Interaction

Source	DF	SS	MS	F	P
Parts	4	0.0000003	0.0000001	3.77795	0.052
Oper	2	0.0000000	0.0000000	0.27733	0.765
Parts * Oper	8	0.0000001	0.0000000	1.05229	0.413
Repeatability	45	0.0000007	0.0000000		
Total	59	0.0000011			

Gage R&R

Source	VarComp	95% Upper Bound	%Contribution (of VarComp)	95% Upper Bound
Total Gage R&R	0.0000000	0.000	80.30	95.15
Repeatability	0.0000000	0.000	80.30	99.53
Reproducibility	0.0000000	0.000	0.00	15.16
Operator	0.0000000	0.000	0.00	15.16
Part-To-Part	0.0000000	0.000	19.70	64.92
Total Variation	0.0000000	0.000	100.00	

Process tolerance = 0.003 in, 76.2 um
Number of Distinct Categories = 1
95% Upper Bound = 1.92406
* WARNING * Probability of part within spec limits = 1.

Figure 15 Gage R&R charts and results.

reading to the next. Double-flank contact has a combination of rolling and sliding action under contact pressure that may further influence these characteristics. Further consideration of the ANOVA method shows that it is very sensitive to changes in the part itself, yet those changes are incorrectly reported as variation from the measuring instrument.

To quantify this effect, a fresh part of the type used in the gage R&R study was rolled 100 consecutive times in double-flank tests by one operator to evaluate total composite error. The results for this plastic acetal gear are shown (Fig. 16).

The upper chart is a chronological run chart reporting total composite error on a clean, un-lubricated plastic acetal gear, with a precision master in successive rolls. The lower chart is a moving range chart where two data points from the upper chart are used to calculate the moving range (for information only). The range data shows that the first four rolls are stable and repeatable, but beyond that the data is not stable. Twenty-four out of 50 sub-groups (48 percent) are out of control on the Xbar chart, which is a conclusive result given that chance alone would have only explained for 0.7 percent of the total sub-groups being out of control. In the first 10 rolls, there is a 30 percent variation in the total composite error result. By roll 42 a noticeable change occurs in the total composite error, and by roll 58 there is a permanent shift and increase in the result. Something is changing with repeated rolls!

This observation does not necessarily mean that the double-flank composite inspection equipment is not repeatable. Since the control chart data is displayed in time order, the conclusion is that the gear has distinctly changed after the fourth roll, continues to change thereafter, and is likely responsible for nearly all of the repeatability error results according to the analysis of variants method.

Hence it can be concluded that 100 consecutive rolls of the same gear on a high-end, well- controlled piece of double-flank inspection equipment were not repeatable since the part itself had changed. It can be further concluded that the gage R&R method in this case does not accurately predict the actual measurement system performance, since the ANOVA mathematics assume that the part does not change.

Further assignable causes of the poor gage R&R. Subjecting gearing to a classical AIAG gage R&R analysis has always been reported to be very difficult, if not impossible. This case study may be the first time in-depth research has been reported to determine why it is so difficult. The following assignable causes may not fully capture all related issues that affect gage R&Rs, but are intended to demonstrate what was observed in this example.

Static measurements**:** Gage R&Rs are most successful with static measurements. An example would be of a gage that measures specific shaft diameters where the accuracy and bias of the instrument is calibrated to a master value on equipment with 10X the accuracy of the measuring instrument to be used.

Dynamic measurement**:** DFCI is a dynamic measurement system and dynamic systems, whether electronic or mechanical (or both), have characteristics that can have a critical impact on the gage R&R results. Some of these characteristics include:

Composite effect of the measurement**:** Rolling a master and a work gear is not just measuring one dimension, even though the result may be expressed as a single value. The result has many influences, often compounding on each other.

Infinite number of readings**:** The composite test result is not a single value but is a dataset made up of potentially hundreds or thousands of data points. Under a dynamic data collection method the algorithms that are used to identify and report on the data may not necessarily provide an identical scan location every time the part is rolled.

Hunting teeth**:** Hunting tooth combinations will ensure that it is unlikely that the same master and test gear flank sets will be repeated on any roll without indexing the roll set every time.

Surface imperfections**:** Surface imperfections will skew the inspection data very quickly, regardless of whether they wear-off or wear-in. The data may show composite error variations roll-to-roll as burrs, flash or irregularities change.

Cleanliness**:** Since the measurements are evaluated in micro-inches or micrometers that are very small units of measure, laboratory control and cleanliness methods are critical. If the debris is moved from one roll to the next, the result will be an inconsistent reading

Running-in/wear**:** DFCI is usually done without lubrication and, depending on the test pressures and materials used, contact may induce run-in or wear characteristics on a gear.

Sliding vs. rolling**:** Minute variations in specific sliding from tooth to tooth may cause variances. Higher sliding action, as is the case in crossed-axis helical gears, may result in slip stick conditions during meshing that could result in more reading instability.

Speed of rolling and data acquisition**:** Higher rolling speeds during testing can impact results — both for mechanical and electronic systems. Rolling speed must take into account the natural frequency response of the tester.

Automated acquisition systems**:** For motorized rolling and computerized, electronic data acquisition, a minimum of 20 data points-per-tooth are recommended. More data points are

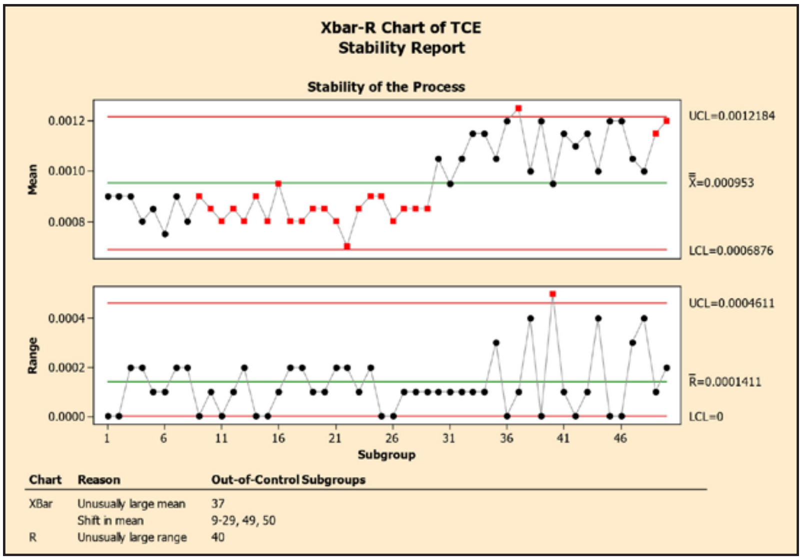


Figure 16 Data control chart of the same acetal plastic gear rolled tested 100 times on the double-flank tester.

always encouraged, but may then produce different results; the computer must be able to process all information at a fast enough speed to obtain an accurate result. In addition, where and how consistently the software and electronics take probe readings is a factor. Variation from where the initial to final points are taken on a specific tooth could impact results.

***Rolling resistance*:** Rolling resistance of various test gears on the mandrel can induce increased torque resistance, resulting in increased gear separation forces that may affect measurement results.

Conclusions related to Gage R&R evaluation in DFCI. Due to the nature of the dynamic response and composite character of TCE, TTE, tight mesh center distance and test radius, an ANOVA analysis of DFCI data may not be appropriate. The sensitivity of the ANOVA mathematics, a requirement for multiple rolls on a single part and all the potential influences of variation, work to defeat this type of analysis. The analysis is often unsuccessful in obtaining a meaningful gage R&R statistic.

Which begs the question: What is the ultimate goal in evaluating the measurement system? The obvious answer: To determine the uncertainty in the measuring system for the purpose of quantifying the true, exact value of the object being measured. Therefore an alternate technique to gage R&R is offered, using certainty analysis for assessing the accuracy of the dynamic DFCI measurement system.

Certainty analysis of the DFCI system. Uncertainty of measurement is defined as: The difference between an actual and the predicted measured value. In understanding measurement systems analysis, the concern is not only with the measured value, but also the error associated with its measurement. Any unknown measurement deviation that departs from the true value is a source of uncertainty. Uncertainty analysis is a predictive technique used to quantify systemic error that is always inherent within the total measurement system.

Building a predictive model of uncertainty for double-flank composite inspection has its own unique forms and practices. The following suggests a practice for calculating uncertainty in DFCI equipment. The output of the uncertainty analysis in the case of DFCI is a linear value (usually in microns) for the specific gage, master gear and part to be measured. The recommended practice is to set an operating tolerance that is inside the required tolerance by at least the amount of the uncertainty result obtained by the following analysis. For example, if a TCE specification is 100 microns and the uncertainty analysis shows a potential of 10 microns of measurement uncertainty, then the design specification should be reduced from 100 to 90 microns in order to meet the original requirement.

The U95 measurement for uncertainty. Equation 25 is the basic form of the U95 measurement for uncertainty as described in Reference 5:

NOTE: All parameters must be in units of like measure (μm or μin)

$$U_{95} = K\sqrt{U_M^2 + U_{R1}^2 + U_A^2 + U_P^2} \tag{25}$$

where:

- U_{95} total uncertainty model with the measuring system and process taken the statistical 95 percent confidence level
- K statistical coverage factor for a specific confidence level
- U_M uncertainty associated with the specific accuracy of the master gear
- U_{R1} uncertainty associated with the repeatability of multiple rolls, usually determined by artifact or alternate process
- U_{R2} uncertainty associated with the system reproducibility element of the measuring system process
- U_A uncertainty associated with the gage blocks or setting discs used to set up tight mesh center distance or test radius according to its calibration accuracy; this uncertainty is not needed for calculations related to total composite error or tooth-to-tooth error, since this additional calibration is not required for that measurement
- U_P uncertainty associated with the probe or instrumentation readout

Further explanation of these factors with a calculation of the same test case as the gage R&R example is detailed in the sections that follow.

Coverage factor, K. The coverage factor is used to expand the uncertainty estimation on the basis of the level of the confidence interval. For a 95 percent confidence interval, a value of $K = 2.0$ is appropriate.

NOTE: K *is derived from the student 't' distribution, as sample size goes to infinity. Within 95 percent confidence intervals 't' converges to 1.960, as described (Ref. 10). For the calculation of uncertainty, the coverage factor is typically rounded up to 2.0.*

Accuracy class of the master — U_M. The accuracy results on the double-flank tester must also account for error in the master gear. Master gears usually come with a certification for either total composite error or run-out error. If the total composite variation of the master is available, its value is to be used in the equation:

$$U_M = \frac{F_{idv3}}{\sqrt{3}} \tag{26}$$

where:

F_{idv3} is the actual total composite variation of the master

If only a run-out certification is available for the master, the equation is modified to adjust for the use of the run-out as follows:

$$U_M = \frac{1.35\,F_{r3}}{\sqrt{3}} \tag{27}$$

where:

F_{r3} is the actual run-out of the master

In this case study, F_{r3} was determined by certification to be $4.4\,\mu m$. Therefore,

$$U_M = \frac{1.35\,F_{r3}}{\sqrt{3}} = \frac{1.35\,(4.4)}{\sqrt{3}} = 3.4295\,\mu m$$

Repeatability uncertainty — U_{R1}. Repeatability is the variation between successive measurements of the same item, taken the same way, under the same conditions. Repeatability is based on a minimum sample of 30 measurements with a dedicated artifact such as:

- Two master gears rolled together
- A master and a test gear rolled together
- An eccentric disk running against a fixed mandrel
- A concentric disk with OD flats running against a fixed mandrel
- Measuring test radius against gage blocks or a calibrated master gage

Further information on artifacts can be found in Reference 6.

In the case of rolling two master gears, the gears shall be tooth indexed, with the teeth re-set to the same positions with every roll.

Multiple rolls of a master gear with a plastic or powder metal test gear usually result in less repeatability than other methods. If this method is chosen, use lubrication between the master and test gear.

The repeatability uncertainty can be estimated as follows:

$$U_{R1} = \frac{\sigma_{30}}{2} \tag{28}$$

where:

σ_{30} is the standard deviation over 30 rolls of the double-flank parameter being evaluated

In this case study, σ_{30} was determined by measurement of center distance on a double-flank gage using gage blocks between mandrels to be 2.0828 μm. Therefore,

$$U_{R1} = \frac{\sigma_{30}}{2} = \frac{2.0828}{2} = 1.0414\,\mu m$$

Reproducibility uncertainty — U_{R2}. Rolling a hardened steel master gear against an unfilled, thermoplastic involute under a tight mesh pressure will induce surface point deflections and flank wear — even at the most minute and microscopic levels. For this reason — and particularly for plastic gearing — it is recommended to use the lowest possible pre-set test force that generates repeatable results.

Uncertainty associated with the reproducibility of the measurement system is a function of repeated inspection equipment set-ups and any bias induced through different operators. The component of reproducibility should be determined with a minimum of three operators and five different parts. Each part with each operator is set-up anew with the inspection equipment and measured only once; one master gear and five test gears are required. The master and each test gear should be indexed to start and end at the same tooth position with every appraiser's roll. The total variation is taken as the mean standard deviation between all parts and appraisers.

The reproducibility uncertainty can be estimated as half of the mean standard deviation between the appraisers as follows:

$$U_{R2} = \frac{\sigma_A + \sigma_B + \sigma_C}{6} \tag{29}$$

where:

σ_A, σ_B, σ_C are the standard deviations for appraisers A, B and C, and the five test parts on the double-flank parameter being evaluated

Using the measurement data from the gage R&R case study —

$$\sigma_A = 3.115\,\mu m$$
$$\sigma_B = 3.117\,\mu m$$
$$\sigma_C = 3.119\,\mu m$$

$$U_{R2} = \frac{\sigma_A + \sigma_B + \sigma_C}{6} = \frac{3.115 + 3.117 + 3.119}{2} = 1.0414\,\mu m$$

Uncertainty of the master gage blocks or setting discs — U_A. When measuring tight mesh center distance or test radius, an additional calibration step must be made between the mandrels that hold the test gear and the master gear. The standard uncertainty of the gage block or setting disc comes from its calibration report, hence:

$$U_A = U_{Gage\ Block\ Calibration\ Error} \tag{30}$$

where:

$U_{Gage\ Block\ Calibration\ Error}$ is the difference between the calibrated value and the nominal value of the gage block or setting disc used

NOTE: For setting discs it may be appropriate to use half the calibrated value if the calibration is based on a diameter and only the radial portion is used.

For total composite and tooth-to-tooth error in this case study —

$$U_A = 0\,\mu m$$

since the gage block was not used for those readings.

For tight mesh center distance or test radius —

$$U_A = 0.0508\,\mu m$$

which was the calibration error associated with the gage block employed.

Uncertainty of the measuring probe or instrumentation readout — U_P. In the case of electronic gages, the transducer usually has a calibration certificate associated with the calibration error of the reading. In the case of a dial indicator, a calibration result would need to be obtained through the gage calibration procedure. The uncertainty factor related to the probe or instrument readout is:

$$U_P = \frac{U_{Probe\ Calibration\ Error}}{\sqrt{3}} \tag{31}$$

For this case study an electronic probe was used with a calibration reading error of 1.05 μm.

Therefore,

$$U_P = \frac{1.05}{\sqrt{3}} = 0.6062\,\mu m$$

Uncertainty case study/numerical example. For the same case study as the previous gage R&R example using Equation 25, the potential variation of the double-flank measurement system (i.e., uncertainty) is:

$$U_{95} = K\sqrt{U_M^2 + U_{R1}^2 + U_A^2 + U_P^2}$$
$$= 2.0\sqrt{3.4295^2 + 1.0414^2 + 1.5685^2 + 0^2 + 0.6062^2} = 7.918\,\mu m$$

For the same case study data as the gage R&R case study example — but with added uncertainty for tight mesh center distance or test radius — the potential uncertainty is:

$$U_{95} = K\sqrt{U_M^2 + U_{R1}^2 + U_A^2 + U_P^2}$$
$$= 2.0\sqrt{3.4295^2 + 1.0414^2 + 1.5685^2 + 0.0508^2 + 0.6062^2} = 7.919\,\mu m$$

The total composite tolerance on the part from the previous gage R&R study is 76.2 μm. The potential uncertainty of our measurement system is approximately 7.92 μm. This means that, for practical purposes, the total composite measurement should be limited to $76.2 - 7.92 = 68.3\,\mu m$ using the uncertainty approach. This is done in order to be confident that true values of measurement will not exceed 76.2 μm. In this case the uncertainty of our measurement system represents 10.4 percent of the total composite error specification.

In contrast, based on the exact same gaging methodology, the gage R&R result would be an unacceptable 80.3 percent.

Uncertainty analysis presents a practical and acceptable alternative for measurement system analysis in dynamic double-flank composite inspection vs. the traditional gage R&R.

Recommendations for Use

At its best, DFCI is used as a screening tool developed for in-process, real-time manufacturing inspection of production gears. Used this way, double-flank composite inspection has, for many years, proven invaluable to conveniently determine changes in gear process and quality elements. Individual gear elemental quality usually cannot be discerned from the results. However, the results can flag a response that indicates there is a change or issue that can be further investigated and resolved. Many gear suppliers and customers successfully rely on a maximum value of total composite error and maximum tooth-to-tooth error to approve or reject a part or lot as an economical form of inspection and validation.

If the value of TCE and TTE is specified intelligently, results in statistical control can be very insightful, relative to gear quality and even performance. When designing gears, proper allowances need to be made for total composite tolerances to ensure that operational backlash will not be compromised. Tight mesh center distance is a powerful way to control functional tooth thickness for backlash control. Although test radius is a specification that has been widely used in the past, tight mesh center distance should always be specified instead of test radius to avoid ambiguity.

Master gear designs should be verified for proper mesh on the gage with the test gear at the maximum and minimum tolerance according to the recommendations in this document.

Statistical methods are a powerful way to assist in the development of tooling for production and for monitoring ongoing production capability in a cost-effective manner. These methods include:

Initial, total composite error probability plots to determine if sample variation is expected or includes special causes when used for evaluating gear tools and dialing in manufacturing processes (Figs. 9a and 9b).

Initial and ongoing total composite error Cpk snapshots to report the strength and robustness of the process, relative to the specified design limits (Fig. 10).

Ongoing probability distribution plots to report predictive values of total composite error, relative to an arbitrary center line. Reporting the percentage of predicted TCE values greater than the middle of the actual design specification gives a good sense and position of where the quality level may be in the future (Fig. 11).

Control charting of tooth-to-tooth error in near-real-time, since it will be the first and most critical flag to reveal if the gear manufacturing process is stable or trending towards increasing or decreasing variation. TTE should be used for predicting and investigating assignable causes of unexpected variation (Fig. 12).

Ongoing capability analysis of tight mesh center distance. Properly interpreted, tight mesh center distance can be a more useful tool in controlling gear quality statistically than total composite error, since it includes the effect of tooth thickness. Adjustments in the ongoing process should be considered when practices as outlined are followed (Figs. 13 and 14).

These methods are insightful at discerning actual and potential gear quality issues that cannot be seen in any other way. This is particularly significant when dealing with high-value-added assemblies.

Although gage R&R is the industry norm for measurement system validation, uncertainty analysis is a better way of dealing with a dynamic measurement system such as double-flank gear inspection.

Although not discussed in this document, further consideration may be made to map double-flank composite inspection charts with in-process flags that help control positive and negative performance issues such as noise, vibration, backlash and design life implications.

This subject may make for a meaningful future discussion opportunity.

References

1. AGMA 930–A05. Calculated Bending Load Capacity of Powder Metallurgy (PM) External Spur Gears; Annex E: Calculations of Inverse Functions for Gear Geometry.
2. Eberle, F.P. "The Real Truth Behind Cpk and Ppk Capability and Potential Process Studies," *Gear Solutions*, October 2012, pp. 40–46.
3. GR&R Definition. *http://www.dmaictools.com/dmaic-measure/grr*.
4. Bower and Touchton. "Evaluating The Usefulness of Data by Gage Repeatability and Reproducibility," *http://www.minitab.com/uploadedFiles/Shared_Resources/Documents/Articles/evaluate_usefulnss_of_data_gage_r_r.pdf.*
5. AGMA ISO 10064–5–A06. Code of Inspection Practice–Part 5: Recommendations Relative to Evaluation of Gear Measuring Instruments.
6. ANSI/AGMA 2116–A05. Evaluation of Double-flank Testers for Radial Composite Measurement of Gears.
7. ANSI/AGMA 2015–2–A06. Accuracy Classification System: Radial Measurements for Cylindrical Gears.

For Related Articles Search

double flank

at www.geartechnology.com

Fred Eberle is a technical specialist in the development of gearing, drive motors and power closure devices in the automotive industry. Over the last 24 years he has worked as a gear development engineer in the mill gearing, lawn & garden and commercial industrial gear industries. He has a bachelor's degree in mechanical engineering from the Rochester Institute of Technology and is a certified Six Sigma Master black belt. Eberle's participation in various AGMA technical committees spans 23 years, having served as chairman of the Powder Metallurgy Committee, and currently on the Plastic and PM Gearing Committees. Eberle has authored several papers on gearing, measurement system analysis and process statistics.

Ernie Reiter, P. Eng. a consultant specializing in the design of gears and geared products, received his degree in mechanical engineering in 1985 from the University of Waterloo in Ontario, Canada. He has authored modern software on gearing and other mechanical components, providing his clients with gearing-related design, consulting, software, gaging, training, and support. Reiter has worked in the field of plastics part production for the automotive industry, specializing in tooling development and directing the manufacture of molded plastic gears. As part of his engineering duties he has acquired advanced skills in computer graphics and their application to gear geometry. Reiter is active in five AGMA technical committees, including vice chair positions in both the Plastics and the Powder Metal Gearing Committees, and is an active participant on the committees for Gear Accuracy, Worm and Fine-Pitch Gearing. Company website: *www.webgearservices.com;* email: *ereiter@webgearservices.com.*