USING TOLERANCE INTERVAL METHOD AS AN ALTERNATE APPROACH FOR MONITORING PROCESS PERFORMANCE (PROCESS CAPABILITY) OF SURFACE ROUGHNESS OF GEAR TOOTH FLANKS TO AVOID GRINDING BURNS

Akaff S. Diam[1]
Julie Z. Zhang, Ph.D., Professor[1]
Ali E. Kashef, Ph.D., Professor[1]
[1]University of Northern Iowa
akaffd@gmail.com; julie.zhang@uni.edu; kashef@uni.edu

Abstract

The process capability measures most commonly referred to as process capability indices (PCIs), C_p, C_{pk}, C_{pm} evaluate process yield, process consistency and statistical variation compared to the specification. This paper explores an alternate approach for monitoring the process performance of tooth flank surface roughness in gear tooth grinding to avoid grinding burn. The gear tooth flanks are susceptible to microstructural damage due to high thermal energy generated during the process of grind operation. The thermal damage in ground gears can cause catastrophic failure in gears' intended life. To avoid grinding burns the process parameters (feed, speed and depth of cut) are selected such that thermal damage should not occur, however, setting up such parameters could affect the process capability of surface roughness (finish) at a desired level. The surface roughness is commonly specified in design as unilateral tolerance and considered as a critical or key characteristic, therefore, most of the automotive or component manufacturing industry requires it to exhibit a process capability index (C_{pk}) of 1.33 or above. Achieving and maintaining process capability of surface roughness is difficult due to the nature of process, moreover, the thermal damage of the ground feature can impair the product life. In this paper, for gears that are made with ferrous materials (plain carbon steels or alloy steels), tolerance interval method is proposed for monitoring the process performance of roughness in gear tooth grinding process as compared to the conventional process capability.

1. Introduction

All processes show variations and the variations in manufacturing processes are due to common causes and special causes. Process capability was introduced in industry to monitor if the common or chance causes of variations as compared to the specification are at a certain level of acceptance. Process capability is evaluated when the process is in statistical control. A process that is operating with only chance causes of variation present is said to be in statistical control. (Montgomery, 1997, p. 130). Several process capability indices (PCIs), C_p, C_{pk}, C_{pm} are available to quantify if the process meets the requirements set by the designers or customers. Mostly C_{pk} index is used which is defined as the minimum of CPU or CPL, where CPU and CPL referred to as upper capability index and lower capability index (AIAG, 2009). Other definition is $C_p = (U - L) / (6\sigma)$ and $C_{pK} = \min \{(U - \mu)/(3\sigma), (\mu - L)/(3\sigma)\}$, where U and L are upper and lower specifications respectivley (Samual &

Norman, 2002). For unilateral tolerances only upper capability index or lower capability index is used whichever is applicable. If the design print calls out maximum tolerance only, CPU is measured. It considers process average and evaluates the process spread with respect to where the process is actually located. The magnitude of C_{pk} relative to C_p is a direct measurement of how away it is from the target.

The assumption for capability index evaluation is that process is approximately normally distributed. If the process variation is centered or targeted between its specification limits, the calculated value for C_{pk} is equal to the calculated value for $C_{p.}$ But with the process variation away from the specification center, the C_{pk} index degrades. Generally, a $C_{pk} \geq 1.33$ indicates that a process is capable. Values less than 1.33 tell that the variation is either too wide compared to the specification or that the location of the variation is off from the center of the specification. It could be the combination of both spread and location measures that how far the process mean is from the nearer specification limit in terms of 3σ distances. C_{pk} works well only for the bell-shaped "normal" (Gaussian) distribution. For others it is an approximation. McCormick (2007) mentioned that knowledge of the mean and variance allows one to calculate with certainty where any value or range of values lies within the distribution. This is not necessarily the case for other distributions. The shapes of non-Normal distributions are usually affected by moments not expressible in terms of the mean and variance alone.

Generally, gears are made of plain carbon steels or alloy steels of various chemical compositions depending upon the application. The choice depends upon number of factors including size, service and design (Dudley, 1984, pp.4.6). There is a wide range of international standards available in industry for ferrous materials used for gear manufacturing. The gears included in this study are subjected to carburizing, quenching and tempering processes to achieve the required surface hardness and relatively softer core hardness for desired mechanical properties during the application and service. The microstructure of heat-treated gears is therefore very important for desired properties.

In gear dimensioning schemes the surface roughness is normally specified as maximum tolerance instead of bilateral tolerance. The ground gears have specification between 0.8 Ra (max) to 1.5 Ra (max) depending upon the type of application of gears. Surface roughness of many precision gears are held at 0.8 Ra max (Dudley, 1984). Since the gears in their application are meshed with the mating gears and rub constantly, so the tooth flank roughness (finish) really matters along with other gears geometrical features and thus regarded as a critical or key characteristic. Gear load capacity is also affected by the gear grades and the surface roughness. The experimental investigations and service experience indicate that a relationship exists between grades of surface texture and aspects of gear load capacity (ISO, 2006).

2. Gear Tooth Flanks Grinding and Grinding Burns

The grinding is a process that shapes the surface by passes with a rotating abrasive wheel (Dudley, 1984, pp. 5.2). Gear grinding is a complex machining process and is affected by several factors. Wang, Wang and Zhou (2005) reported that grinding is most used process for obtaining high level of surface quality, it remains as one of the most difficult and least understood processes. Apart from achieving surface roughness, the contact between the grinding wheel grain and workpiece material generate a thermo-mechanical load on the workpiece. The most significant disadvantage of increasing the circumferential speed is the rising thermal load acting on the surface layer of the workpiece (Fritz, Sebastian, Patrick, 2016). Some of the important factors for machining process are infeed, wheel speed, depth of cut, types of grinding wheel used and the dressing frequency. The gear grind operation is usually conducted after the heat treatment of semi-finished gears to obtain required sizes and tolerances. A grind stock is left in pre-heat treatment operations of gears' tooth flanks, so that finished operation makes it to required geometrical sizes after grinding. Grinding process generates high thermal energy and if the parameters are not set correctly, could damage the microstructure of tooth flanks. The inspection of thermal damage is done with an acid etch process. Grinder burn is a broad term and encompasses varying degrees of thermal damages. To visually detect, any burn must exhibit enough contrast between grey and dark grey (Crow & Pershing, 2018). Gears having thermal damages could adversely affect the product life and cause premature failure. Grinding speed, and depth of cut could create thermal damage. According to Augier, Cruz, Paula and Bianchi (2008), the measurements of surface roughness show an increase in magnitude as the depth of cut was increased, mainly after the test with 35µm depth of cut where grinding burn on the workpiece surface took place. Controlling the grinding parameters is a key to avoid grinding burns, which however could impact surface roughness process mean that could potentially degrade the process capability (≥1.33) of surface roughness. Some of the factors causing grinding burns are:

- Grind stock amount on the tooth flanks – overstock / understock
- Pitch diameter runout in the semi-finished gears
- Gear geometric features' (gear lead, gear profile and tooth thickness) variations
- Heat treatment distortion
- Material variation
- Oil nozzle placement
- Machine set up

The grinding process must be optimized to a level to avoid thermal damages. This situation, however, causes the roughness measurements skewed and most of the data concentrate close to the upper / maximum tolerance level, creating a non-centered process for surface roughness values. The underlying assumption for process capability analysis is that data should exhibit a normal distribution beside being statistically predictable and under control (AIAG, 2005). Hence, achieving capability of e.g. C_{pk} of 1.33 becomes difficult. Moreover, in batch processes, batch to batch variation is inherent and mostly dependent on how well the process is controlled in heat treatment and pre-heat treatment manufacturing processes of steel grades used . Post-heat treatment processes are greatly affected by the consistency of carburized area of the gear tooth flanks. During heat treatment process various factors affect the gears distortion. According to Davis (2005), in gears two types of distortion occurs, one is body distortion, which includes run-out, out-of-roundness, out-of-flatness. The second is the distortion in gear tooth geometry. Body distortion also influences the tooth geometry distortion a great deal. The grinding stock is thus ensured and is determined on the basis of cleaning of all the surfaces of teeth considering distortion and growth of gears after carburizing and hardening (Davis, 2005). Also, when tool wear and adjustments are frequently made, typically maintaining process mean for surface roughness at certain level is difficult. So even if a batch displayed surface roughness process capability of 1.33 or above, there will be no surety that process will always be running at or above 1.33 C_{pk} index in the next batch. Achieving and monitoring capability index of 1.33 and above is therefore has ever been challenging for surface roughness in gear grinding while ensuring no grinder burn should happen.

3. Statistical Tolerance Interval, an Alternate Approach

A statistical tolerance interval is an interval that one can claim to contain at least a specified proportion, *β (also P is used in some texts)*, of the distribution with a specified degree of confidence, 100 (1-α) %. Such an interval would be of particular interest in setting limits on the process capability of a product manufactured in large quantities (Hahn, Meeker & Escobar, 2017). According to the international standard a statistical tolerance interval is an estimate interval, based on a sample, which can be asserted with confidence level (1- α), for example 0.95, to contain at least specified proportion p of the items in the populations (ISO, 2014). The international standard discusses two methods for determining the statistical tolerance intervals, a parametric method for the case where the characteristic being studied has normal distribution and a distribution-free method for the case where nothing is known about the distribution except that it is continuous. Before using tolerance interval that depends heavily on the normality assumption, one should assess the adequacy of the normal distribution as a model, (Hahn & Meeker, 1991). As reported by Hahn and Meeker, two-sided distribution-free statistical intervals from a random sample from a specified population one generally proceeds as follows:

- Specify the desired confidence level for the interval
- Determine (from tabulations and calculations) the order statistics (ordered observations from smallest to largest are called ordered statistics) that provide the statistical interval with at least the desired confidence level for the sample size
- Use the selected order statistics as the end points of the distribution-free interval

One sided distribution -free bounds are obtained in a similar manner, except that only the order statistic is used as the desired lower or upper bound (Hahn & Meeker, 1991). Tolerance interval method is useful in one sided tolerance (upper bound or lower bound) where process centering (targeting) is an issue and is less important than feature's conformance. Also, non-parametric (distribution-free) tolerance intervals methods can be used if the measured data do not follow normal distribution to assess the conforming proportions.

4. Statistical Tolerance Interval Procedures

There are several procedures available in text to construct the statistical tolerance intervals depending upon the requirements. This paper will discuss following statistical tolerance interval procedures.

4.1 Upper-bound tolerance interval when the population is normal distribution (with known mean and known variance)

The international standard (ISO, 2014) provides procedure when the values of the mean, μ, and the variance, σ2, of normally distributed population are known, the distribution of the characteristic under investigation is fully determined. There is exactly a proportion P of the population to the left of of the one-sided interval equation below;

$$XU = \mu + \mu_p \times \sigma \qquad [1]$$

In this equation, μp, is the P-fractile of the standardized normal distribution, XU is the upper limit of statistical tolerance interval, σ is the population standard deviation and μ is the population mean.

4.2 Upper-bound tolerance interval when the population is normal distribution (with unknown mean and known variance)

If we have a normal population with unknown mean and known variance, find k such that $\bar{x}+k\,\sigma$ satisfies that *at least* a proportion P of the population is below $\bar{x}+k\,\sigma$. Note that $\mu + \mu_p\,\sigma$ is the population tolerance limit in the sense that exactly a proportion P of the population is below that limit. So if $\bar{x}+k\,\sigma \geq \mu+\mu_p\,\sigma$, where $\bar{x}$ is the sample mean and k is the factor used to determine the tolerance limits then the proportion of the population that is smaller than $\bar{x}+k\,\sigma$ is at least P. Thus the probability (confidence level) that a proportion of the population is at least P is (1-α), if

$$P\ [\bar{x}+k\ \sigma \geq \mu + \mu_p\, \sigma] = 1\text{-}\alpha \qquad [2]$$

The probability on the left side of [1] can be rewritten:

$$P\ [\bar{x}+k\ \sigma \geq \mu + \mu_p\, \sigma] = P\ [(\sqrt{n}(\bar{x} - \mu)/\sigma \geq \sqrt{n}\mu_p - \sqrt{n}k] = 1\text{-}\alpha \qquad [3]$$

The variable $\sqrt{n}(\bar{x} - \mu)\ /\sigma$ in formula [3] has a standard normal distribution, and follows from the last equality in [2] that:

$$\sqrt{n}\mu_p - \sqrt{n}k = \mu_\alpha,$$

Which can be rewritten as:

$$k = \frac{1}{\sqrt{n}}\, \mu_{1\text{-}\alpha} + \mu_p \qquad [4]$$

The value of k is based on the confidence level (1- α), minimum proportion of the population asserted to be lying in the statistical tolerance P and number of observations in the samples n.

4.3 The distribution-free statistical tolerance interval for any type of distribution

For establishing distribution-free tolerance intervals the order statistics (of sample is determined by solving binomial distribution function for smallest sample size n. International standard (ISO, 2014) describes the procedure for constructing distribution-free or for any type of distribution statistical tolerance bounds. Assume we have samples, $x_1, x_2, x_3, \ldots\ldots, x_n$, of independent random observations on a population (continuous, discrete, or mixed) and let its order statistics ((ordered observations from smallest to largest are called ordered statistics) is $x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq x_{(4)} \leq \ldots\ldots\ldots \leq x_{(n)}$. The interval with 100(1-α) % confidence that at least 100 P % of the population lie between the v^{th} smallest observation and w^{th} largest observation is:

$$\sum_{x=0}^{v+w-1} \binom{n}{x} p^{n-k} (1-p)^x\ \leq \alpha, \qquad [5]$$

where v ≥ 0, w ≥ 0, v + w ≥ 1, 0 < α < 1.

Hanson and Owen discussed that continuity requirement on cumulative distribution function (c.d.f.)'s is unnecessary. They noted that many (c.d.f's) which occur in practice are not continuous, and in many cases where distribution-free tolerance limits are applicable they are not being used because of an uncertainty as to whether the underlying distribution is or is not continuous or because of the certainty that it is not (Hanson & Owen, 1963). So, when the (c.d.f.) of the population characteristic X is not continuous, the statement will be slightly modified such that there is at least 100(1-α) % confidence that at least 100 P % of the population is between $x(_v)$ and $x(_{n-w+1})$ or equal to $x(_v)$ and $x(_{n-w+1})$. When v + w = 1, formula [4] reduces to:

$$P^n \leq \alpha, \text{ or}$$

$$n = \log(1\text{-}\alpha) / \log(P) \qquad [6]$$

The equation [6] shows the minimum sample size required for certain confidence level and proportion to be in conformance. Thus, to be 95% confident that at least 95% percent population lies below the largest value of the sample, the sample size must be n = log (1-0.95)/log (0.95) = (-1.3) / (-0.022) = 59. There are several tables available which gives minimum samples required for constructing statistical interval at certain proportion p and certain confidence level (1-α)%. Hahn and Meeker have compiled and listed tables based on the selection of sample size n, confidence levels 100(1-α) % and at various levels of conformance proportion percentage P (Hahn & Meeker, 1991). Also, now-a-days with the help of latest statistical analysis softwares like Minitab® it is easily possible to calculate, and construct distribution based or distribution-free statistical intervals.

5. Data Analysis and Interpretation

In this paper, a real-life sample data of 7 different gears (total 2039) ran on the same gear grinder is used and analyzed to calculate one sided (upper bound) tolerance interval. All parts are machined from alloy steel forgings, for example grades 8620, 4120, 20MnCr5 (mainly carbon contents between 0.18%-0.2% along with varying composition of Ni, Cr, Mn, Si, P, S and Mo) and have print tolerance of roughness 1.0 RA max., however the sizes of parts are different in gear geometry, module size and tooth flank face width. Table 1 shows the quantity of each part and the minimum and maximum values recorded during 3 shift operations. Same instrument, Mitutoyo Surface Tester, used for measuring the roughness values for all parts.

Table 1 below shows none of the measured value exceeded the specification limit of 1.0 max, however, the distribution fit for normality failed for 5 out of 7 parts.

Table 1. Statistical Summary of Roughness (Ra) by Part Numbers

Part	Pieces Checked	Roughness Min (Ra)	Roughness Max (Ra)	Roughness Mean *(Ra)*	Std. Dev of Roughness *(Ra)*	Distribution Fit
A	210	0.386	0.977	0.703	0.1251	Normal
B	219	0.451	0.986	0.732	0.1147	Non-Normal
C	381	0.423	0.984	0.714	0.1145	Normal
D	175	0.442	0.998	0.765	0.1291	Non-Normal
E	581	0.410	0.998	0.780	0.1272	Non-Normal
F	306	0.425	0.997	0.778	0.1068	Non-Normal
G	167	0.334	0.998	0.722	0.1592	Non-Normal

There are several reasons that the data failed to be normally distributed. Some of the reasons are as follows:

- Shift to shift variations
- Batch to batch variations
- Operator to operator variations

Running analyses at 95% percent confidence level (1-α) for all seven different parts included in the study for estimating that 98% of population (*P* %) being in tolerance bound. Results are summarized in Table 2, since part A and part C follow the normal distribution the upper tolerance bound for part A and C is 0.988 and 0.968 respectively. This indicates that at 95% confidence, 98% of the population will have values equal to or less than 0.988 Ra and 0.968 Ra for part A and part C respectively, which is less than 1.0 max Ra as specified in the print.

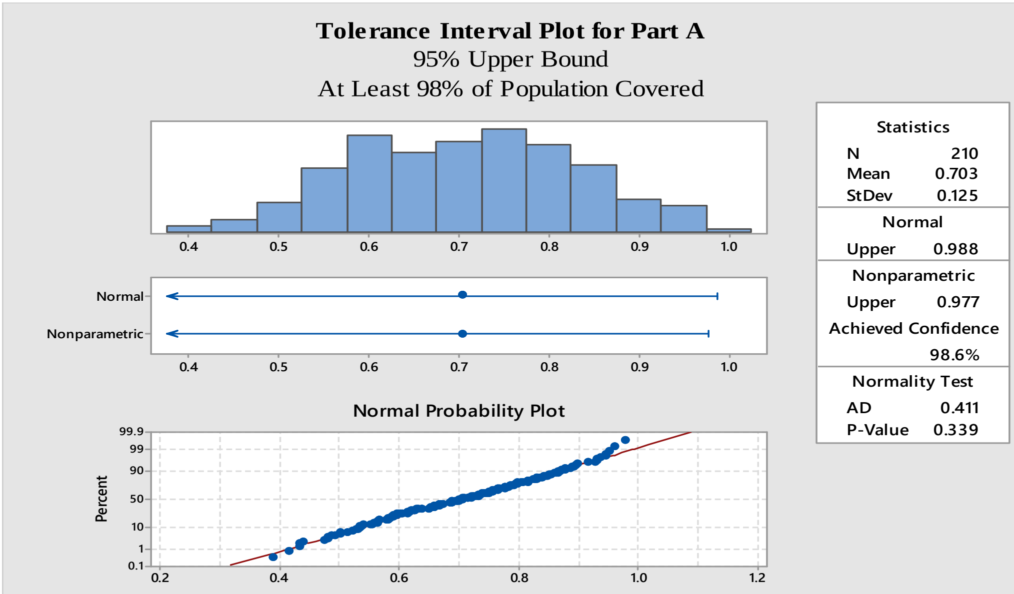

Figure 1. Tolerance Interval Plot for Part A Using Minitab®

Table 2. ***Summary of Calculated Tolerance Intervals of parts***

95% Upper Tolerance Bound			
Part	**Normal Method**	**Nonparametric Method**	**Achieved Confidence**
A	0.988	0.977	98.60%
B	0.992	0.986	98.80%
C	0.968	0.981	98.20%
D	1.062	0.998	97.10%
E	1.057	0.991	97.50%
F	1.017	0.993	98.50%
G	1.089	0.998	96.60%

Achieved confidence level applies only to nonparametric method.

Table 1 shows that the sample data for parts B, D, E, F and G do not follow the normal distribution, therefore, nonparametric method is used, the results indicated in the column "Achieved Confidence" in Table 2 are all greater than 95%.

Now we can analyze the process capability of the same sample and compare the results achieved in tolerance interval methods. Part A and C capability indices with upper and lower confidence interval bounds on the capability indices are given in the Table 3. The sigma used for calculating the process indices is the overall process sigma for each part. Table 3 showed that the capability indices for part A and C are less than 1.0, which means the process is not capable.

Table 3. ***Process Capability of Normally Distributed Parts***

Part	C_{PK} (C_{PU})	Confidence Interval on C_{PK}		Projected Performance
		Lower CI	Upper CI	PPM (percent)
A	0.79	0.70	0.88	8,933 (0.89%)
C	0.83	0.76	0.90	6,360 (0.63%)

To improve the process capability, process variation must be minimized. Figure 2 shows that reason for the part A process being not capable is that the mean is not centered.

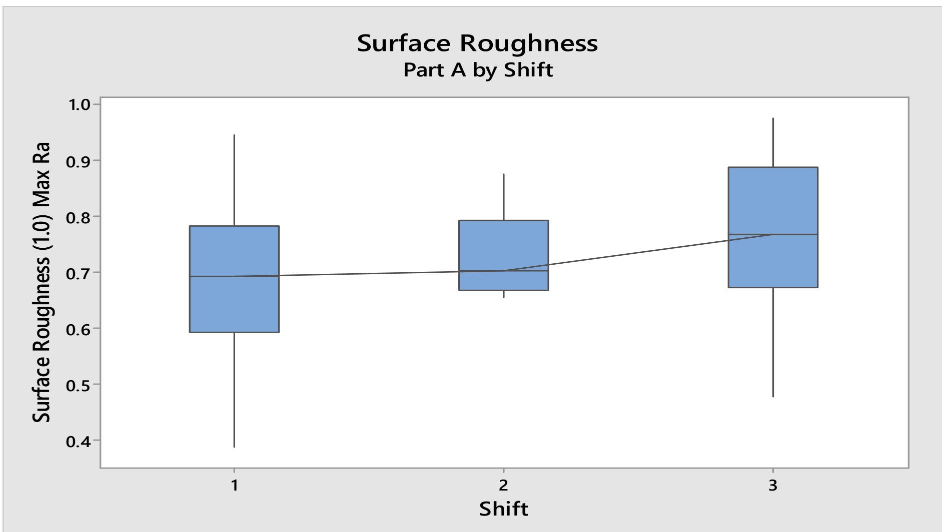

Figure 2

There is difference in means for each operator /shift and for each batch. Figure 2 shows the shift to shift difference of data in box plot for Part A. Actions must be taken to control the machine performance, consistency of input material, change the basic method by which the process operates or shift the process average closer to target. Frequent process adjustments, sorting and scrap costs will have to be increased to bring the process to an acceptable level of ≥ 1.33. Comparing with the process capability evaluation, tolerance interval method is more practical and economical to monitor the performance of process without sacrificing much for cost and it is providing greater than 95% confidence that 98.6% and 98.2% populations of these parts A and C respectively are conforming to the specification.

6. Conclusion and Further Recommendations

The above results for individual gears made of alloy steels suggest that estimating the tolerance interval has an advantage and provide a confidence that more than 98% population is within tolerance specification. Adjusting unreasonably the process to control the roughness measure and process mean could create issues of grinder burn and could increase the cost of process, cycle time and scrap cost. Companies can use this alternate approach of tolerance interval method where the process mean is not centered, and conventional process capability achievement is difficult. Parts are produced in short batches, parameters for grinding are set by individual operators, and the distribution of roughness measure is concentrated mostly between 0.65> Ra >0. 87. Selecting the gear grinding parameters like feed, speed and depth of cut to bring the mean of surface roughness measure close to the nominal or target will increase frequency of grinding parameter adjustments and hence will also increase the part processing cycle time and the risk of tooth flank grinding burn. Companies can formulate data collection schemes to collect surface roughness measurement data and keep monitoring and recording the process at the start and at the end of shift or with certain cadence between shift. Enough data can be collected over a period, it will become easier and economical way of monitoring the process performance of surface roughness in terms of conformance with a required degree of confidence.

For future studies, processes like reaming, hardness check (heat treatment) or broached diameters can be studied and evaluated if the process capability is mandated. These processes are also targeted where conventional process capability analysis could fail, for example, hardness is specified generally as unilateral tolerance (HRC 58Min) and making parts harder would affect finishing processes in terms of increased tooling costs. Similarly, broaching tools and reaming tools are always selected so that maximum tool life could be utilized, hence the process targeting is non-centered compared to the nominal when the tools are new in the early production of parts with these tools.

7. References

AIAG, 2005. Statistical Process Control (SPC), Second Ed. Automotive Industry Action Group, Detroit, MI.

Augier, P. R., Cruz, E. D. C., Paula, W. C. F., & Bianchi, C. E. (2008). Predicting Surface Roughness in Grinding using Neural Networks Advances in Robotics. Automation and Control. CC BY-NC-SA, 3.

Crow Jonathan R. & Pershing Michael A., Standard Samples for Grinder Burn Etch Testing, Gear Technology Magazine, June 2018, 54-55, Retrieved from https://www.geartechnology.com/articles/0618/Standard_Samples_for_Grinder_Burn_Etch_Testing/

Davis, J. R. (2005). Gear materials, properties, and manufacture. Retrieved from https://ebookcentral-proquest-com.proxy.lib.uni.edu

Dudley, Darle W. (1984). Handbook of Practical Gear Design (Rev. 1st Ed.). New York: McGraw-Hill.

D. W. McCormack, Jr., Ian R. Harris, Arnon M. Hurwitz & Patrick D. Spagon (2000) Capability indices for non-Normal data, Quality Engineering, 12:4, 489-495, DOI: 10.1080/08982110008962614

Fritz Klocke, Sebastian Barth & Patrick Mattfeld (2016) High Performance Grinding, Procedia CIRP, 4, 266-271, doi:10.1016/j.procir.2016.04.067

Hahn, G. J., Meeker, W. Q., & Escobar, L. A. (2017). Statistical intervals: A guide for practitioners and researchers. Retrieved from https://ebookcentral-proquest-com.proxy.lib.uni.edu

Hahn, G., & Meeker, W. (1991). Statistical intervals: A guide for practitioners (Wiley series in probability and mathematical statistics. Applied probability and statistics). New York: Wiley.

Hanson, D., & Owen, D. (1963). Distribution-Free Tolerance Limits Elimination of the Requirement that Cumulative Distribution Functions be Continuous. Technometrics, 5(4), 518-522.

International Organization for Standardization. (2014). Statistical interpretation of data – part 6: Determination of statistical tolerance intervals, ISO 16269-6:2014. Retrieved from https://asc.ansi.org

International Organization for Standardization. (2006). Technical Report, Cylindrical gears – code of inspection practice, ISO/TR 10064-4/:1998. Retrieved from https://asc.ansi.org

Montgomery, D. C. (1997). Introduction to statistical quality control (3rd Ed.). New York: John Wiley & Sons.

Samuel Kotz & Norman L. Johnson (2002) Process Capability Indices—A Review, 1992–2000, Journal of Quality Technology, 34:1, 2-19, DOI:10.1080/00224065.2002.11980119

Wang, J. Z., Wang, L. S., Li, G. F. & Zhou G. H. (2005). Prediction of surface roughness in cylindrical traverse grinding based on ALS algorithm, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, 18-21 August 2005, 549-554, Guangzhou.