Applications of Gauge Reproducibility and Repeatability (GR&R) in Final Finish Tire Testing

Shaun M. Immel, Ph.D., P.E.

Measurement Process Consistency

One particular tool used in the industry for many years to assess measurement instrument repeatability and consistency across several inspection agents is the Gauge Reproducibility & Repeatability (GR&R) tool. This paper will investigate and discuss the specific application of the GR&R tool to evaluate tire final finish machine measurement consistency. As it turns out, there are some issues with how most tire companies use the GR&R tool and special consideration should be taken into account when applying these tools to final finish tire measurements.

Final finish tire testing systems are used around the world in tire manufacturing plants to assess the final quality of many different types of tires. When decisions are made to send tires on to a customer or scrap/downgrade a tire, it directly affects the financial condition of the tire manufacturer. There are both hard and soft costs associated with scrapping or downgrading a tire, especially if this decision is made incorrectly. There are even more extensive costs incurred if a poor quality tire is sent to a customer incorrectly. Typically a tire is measured and compared to a screening limit to make this financial decision. Hence, the tire manufacturer needs to be certain that the quality measurement process is both accurate and repeatable.

The average tire quality measurement must be "correct" when compared to some measurement standard; this is known as measurement accuracy. In the case of tire quality measurement, accuracy is difficult to assess due to the absence of a "NIST[1]-traceable" standard. Measurement accuracy is a separate but related topic and will not be discussed in this paper. Repeatability is typically used to describe a measurement process' ability to return the same measurement over and over again, each time a production piece is measured. The more repeatable a machine is, the less chance there will be in making an incorrect tire quality assessment resulting in either scrapping/downgrading an acceptable tire or even worse, sending a tire with questionable quality to a customer and leading to extensive costs for the tire manufacturer. If a measurement process is not repeatable, then depending on the time a tire is measured by a machine, a different quality decision may be made, particularly for tires close to the screening limit. So knowing the consistency of measurement is extremely important, but what is really required to do this?

The consistency of the measurement process as it applies to final finish tire quality measurement is comprised of three different component effects:

- Machine effects on measurement consistency
- Tire effects on measurement consistency
- Machine/Tire interaction effects on measurement consistency

When assessing measurement consistency, it makes the most sense to isolate machine effects first, which requires special attention to the testing/assessment process. This specific machine evaluation process will be the focus of this paper. It must be done while minimizing tire effects and machine/tire interactions and is done through the intelligent choice of "master" tires. Master tires are tires set aside from the normal production flow, which typically have been selected for their inherent measurement stability and have been exercised for several cycles on separate equipment to further stabilize their measurements. The difficulty is that the properties and

---

[1] National Institute of Standards and Technology, United States Department of Commerce

measurements of these master tires change over time as the tires age and wear as a result of the measurement process itself. The use of these tires and other testing practices including but not limited to hand lubrication is critical to making a proper assessment of the machine's ability to provide consistent measurements.

Once it is known that the machine is measuring consistently, then one can begin to investigate the tire and machine/tire interaction contributions to the consistency of the whole measurement process. After all, the final finish equipment is used to determine the quality of standard production tires, so it only makes sense to understand the consistency of the whole measurement process. This is particularly challenging since production tire measurements can change each time they are measured, particularly in the first few measurement cycles. This analysis effort should only be undertaken after a proper machine consistency evaluation has been successfully completed with master tires.

Regardless of the metric used to assess measurement process consistency, it becomes very difficult to answer the typically posed question of, "is the measurement process consistent?" It is difficult to set limits on any numerical calculation in order to make a final decision. General guidelines for doing this will also be discussed in this paper.

Gauge Reproducibility & Repeatability (GR&R) Overview

There are two primary components of interest related to measurement variation, which the GR&R analysis calculates. Those components are reproducibility (historically intended to measure consistency across several operators – a measurement "person" effect) and repeatability (measure of consistency for a given measurement device – a measurement "device" effect). In the application of GR&R to automated final finish equipment, there is typically no "operator" involved since the equipment is automated; however, there is a potential novel use of this operator component in assessing the agreement (or correlation) of automated tire measurement equipment. The GR&R process can be used to measure how closely two or more automated machines are measuring (as opposed to several operators). The same basic GR&R analysis results interpretation applies, but the reproducibility variable can be interpreted as a level of measurement agreement between individual machines. The GR&R analysis can also be performed on a single machine to determine only its repeatability characteristics by simply ignoring the reproducibility element.

There are two basic calculation methodologies of GR&R analysis in use today

- Average Range Method
- ANalysis Of VAriance (ANOVA) method

The original GR&R method proposed by the Automotive Industry Action Group[2] (AIAG), the most popular automotive industry technique today, uses the Average Range Method. However, the ANOVA-based method seems to be accepted as well and is generally the preferred method. A detailed and thorough review of the exact steps and calculations in a GR&R will not be presented in this paper as there are several great published sources for information on this topic including but not limited to (Automotive Industry Action Group, Automotive Division, 2002),

---

[2] www.aiag.org

(Wheeler, 2009), and (Minitab Inc., 2010). However, a quick review of key indicators of the GR&R will be discussed.

Consider the partial GR&R results in Table 1 obtained by putting example data through an ANOVA-based crossed GR&R analysis in Minitab®, which is a statistical tool that allows users to analyze numerous types of data (Minitab Company Information, 2010). This data represents a GR&R executed with several tires, measured several times, through a few machines. The exact quantity of each of these components is not important to the discussion that follows the example data.

Two-Way ANOVA Table With Interaction

Source	**DF**	**SS**	**MS**	**F**	**P**
Tire No.	14	8145437	581817	223.057	**0.000**
Machine	1	3038	3083	1.182	**0.295**
Tire No.*Machine	14	36517	2608	5.605	**0.000**
Repeatability	120	55840	465		
Total	149	8240877			

Table 1 – Example Gage R&R Study – ANOVA Method Results

Determining the statistical significance of individual factors in the ANOVA section of the GR&R analysis typically requires a probability value (P-value) less that 0.05 (5%). That is, if the P-value is less than 0.05 it is considered to be a statistically significant parameter effecting a change in tire measurements. This 5% limit is generally accepted across the industry; however, there is no discussion of P-values in the AIAG guidelines. Under the Probability (P) column in Table 1, there are two P-values shown to be 0.000. These two values indicate that the measurement values of the individual tires themselves and the interaction between the individual tires and the machine are statistically significant. This is expected, since the hope is that the measurement machines are consistent enough to distinguish the measurements of each tire and the dominant component of measurement variations is the tires themselves having little to do with the measurement machines. The P-value of the machine factor, which contains both a reproducibility and repeatability factor, is 0.295 indicating that it is not known to be a significant factor driving differences in measurement values. In other words, there is NO statistically significant difference between the measurements taken across a few machines (reproducibility) or within the measurements for each tire on a given machine (repeatability) when compared to the variation in the tire measurements themselves. These results depend on the values of the master tires chosen for the GR&R study and merits deeper discussion.

By looking at more GR&R results from this same example data as shown in Table 2, it can be seen in the % Contribution column that the different machines contributes only 0.01% of the total variation in the experiment resulting in a total reproducibility contribution of 0.74%. The repeatability contribution is 0.79% resulting in a total machine contribution (Total Gage R&R) of 1.53%.

Source	**VarComp**	**%Contribution (of VarComp)**
Total Gage R&R	900.3	**1.53**
Repeatability	465.3	**0.79**
Reproducibility	434.9	**0.74**
Machine	6.3	**0.01**
Machine*Tire No.	428.6	0.73
Part-To-Part	57920.9	98.47
Total Variation	58821.1	100.00

Table 2 – GR&R % Contribution Results

The AIAG GR&R method does not specify guidelines for these % Contribution values; however, since they are variances, it is common to consider values <1% generally acceptable, 1% - 9% marginally acceptable, and >9% unacceptable. These values are derived by squaring the AIAG guideline values discussed later in this paper.

Two other results that can be calculated are the % Study Variation and % Tolerance Variation. These values are shown for the example data in Table 3.

Source	**StdDev (SD)**	**Study Var (6 * SD)**	**%Study Var (% Study Var.)**	**%Tolerance (% Tol. Var.)**
Total Gage R&R	30.004	180.03	**12.37**	**18.00**
Repeatability	21.572	129.43	8.89	12.94
Reproducibility	20.885	125.13	8.60	12.51
Machine	2.515	15.09	1.04	1.51
Machine*Tire No.	20.703	124.22	8.54	12.42
Part-To-Part	240.668	1444.01	**99.23**	**144.40**
Total Variation	242.531	1455.18	**100.00**	**145.52**

Table 3 – GR&R % Study Variation and % Tolerance Variation

The AIAG publishes general guidelines for acceptable limits of both % Study Variation (to assess usefulness of measurement device for process improvement) and % Tolerance Variation (to assess usefulness of measurement device as a screening tool). These guidelines state that values of Total Gage R&R less than 10% are considered excellent, between 10% and 30% is acceptable under certain circumstances, and above 30% is unacceptable. In the example data show in Table 3, the Total Gage R&R % Study Variation is 12.37% and the % Tolerance Variation is 18.00%. These values would officially classify this measurement device as marginal but acceptable for both purposes of process improvement and part screening. The reader is cautioned, however, that the AIAG guidelines themselves state that these are “rule of thumb” criteria and that final measurement device acceptance should not come down to a single set of indices, but should include other indices as well as long term monitoring and reviewing of measurement device performance over time. In addition, there are known flaws in this AIAG calculation methodology and another special consideration that should be taken into account when analyzing final finish testing data. These calculation flaws and final finish testing application considerations will be discussed in the following two sections.

Flaws in the AIAG GR&R Process

The AIAG guidelines for GR&R study are known in several circles to be overly conservative and can lead to unnecessary disqualification of measurement processes or devices. This, in turn, can result in unnecessary expenditures for a tire manufacturer to improve a measurement system or device. The inconsistencies occur during the calculation of the % Study Variation and % Tolerance Variation ratios. Notice from the data in Table 3 that the % Study Variation component of Total Gage R&R is 12.37% and Part-to-Part component is 99.23%. These numbers are the two components that should make up 100% of the variation; however, they don't add up to 100%. This is caused by the fact that standard deviations are used in the calculations of these ratios, which cannot occur without violating proper statistical procedure. The ratios are essentially physically meaningless. The same is true for the % Tolerance Variation numbers as well. A much more comprehensive analysis of these mathematical flaws as well as the proposal for a modified GR&R assessment process is presented in (Wheeler, 2009), and for the sake of brevity will not be reiterated in this paper. It suffices to say that serious consideration should be given to adopting Wheeler's method or a similarly improved technique if flexibility exists in this choice for the tire manufacturer. There is an additional special consideration worth reviewing when the AIAG GR&R procedure is applied to the evaluation of tire final finish testing measurement data.

Special Consideration For GR&R Application in Final Finish Testing

There is another significant factor of concern with the % Tolerance calculation as part of the GR&R evaluation process. All measurement data from the tire final finish area (i.e. force uniformity, dynamic balance, and geometry), with the exception of conicity, is positive in value. In addition, these measurements are compared to a single value for the purpose of quality screening. If a measurement is less than the screening (specification) value, it is considered acceptable as a particular class or grade of tire. Conversely, if the measurement is above this screening value it is considered unacceptable for the given quality class under consideration. This situation is described as a one-sided specification scenario. When the % Tolerance parameter is calculated, by definition, it divides the variance (actually standard deviation) of the component under consideration by the range of acceptable product. In the case of a one-sided specification, there is no official "range" of acceptable product.

There are two generally accepted methods to continue with the calculation of the % Tolerance. The first method assumes the lower specification limit to be zero, therefore calculating the acceptable product range as the upper specification limit - zero, which simply equals the specification limit. The % Tolerance parameter is then calculated as 6 times the particular variation under consideration divided by the specification limit. The second method calculates the acceptable product range as the difference between the specification limit and the study mean (assuming the study mean is below the specification limit, which is usually the case). The % Tolerance parameter is then calculated as 3 times the variation under consideration divided by the difference between the specification limit and the study mean. In the author's opinion, there is a strong case to steer clear of this second method since it is heavily dependent upon the mean of the master tires. Even though master tires are generally chosen to represent the anticipated distribution of production tire measurements, it seems unreasonable to condemn a measurement device purely on the average value of master tires. It seems more natural, since an appropriate target for tire production final finish measurements is zero, that this value (zero) should be used as a lower specification limit. The second method does provide some useful information regarding how effective the current measurement system is as a screening tool. If this calculation results in a large value, the tire manufacturer may want to consider using a guard-banded (or otherwise known as an undercut) specification limit to help reduce the probability of unacceptable tires getting shipped to customers, until production yields can be improved by driving the mean away from the specification limit.

Recommendations

This paper has described several steps that a tire manufacturer must take to ensure that final finish measurement machines currently being utilized are measuring tires with the consistency necessary to correctly assess tires. Wrongly assessing tires can lead to substantial financial loss for tire manufacturers who do not take the necessary steps to ensure that their machines and processes are in fact, correctly and consistently measuring tires.

The first step that needs to be taken to address these issues is to isolate the machine effects from the tire and machine/tire interactions. Special attention needs to be given to the machine as well as the master tires to ensure that they (separately and together) are in the correct condition to produce the required results. These “master” tires need to be constantly checked due to the fact that they may change over time due to external factors such as rubber aging and wear and tear from multiple measurement cycles. Exposure of the master tires to intense temperatures and abhorrent storage conditions may cause them to change as well. Special care, such as hand lubrication of tires, needs to be taken into account as well for the machine and tires during the GR&R tests as discussed in earlier sections.

Carefully choosing master tires is a crucial step before performing a GR&R study on a final finish tire testing system. These master tires should represent the distribution of tires that the manufacturing facility expects in the normal production process. Consequently, the average value of the master tires should equal the normal production average value and the distribution of tire values similar.

As previously stated, the AIAG guidelines for GR&R study are known to be overly conservative and can lead to unnecessary disqualification of measurement processes or devices. In many instances, the AIAG method may be the only option for tire manufacturers, so it is very important to understand that this method has flaws. Tire manufactures must be able to recognize and discuss these flaws with their customers and be able to make decisions and corrections to offset the discrepancies with the AIAG method.

The last important point when evaluating GR&R is with the calculation of the % Study Variation and %Tolerance. If the AIAG GR&R method was historically followed, serious consideration should be given to using an improved GR&R methodology if possible. At a minimum, it is in the author's opinion to assume the lower specification limit to be zero when calculating the % Tolerance parameter. If this is not an option, then the effects of master tire measurement distribution on calculating the final finish testing equipment acceptance ratios should be taken into consideration. This information should be used only in connection with other evaluation indices. This will allow the tire manufacturer much more accurate GR&R calculations and ultimately leading to a higher yield of acceptable tires to be passed along to the end user.

This in-depth look into the steps needed to correctly assess the applications of Gauge Reproducibility and Repeatability for a tire manufacturing facility shows us that these decisions affect more than the manufacturing facility itself. These decisions require sound planning and collaboration with suppliers that know this tool best. Ensuring that tires delivered to the marketplace are what they claim to be is the goal of every tire manufacturer. Verification of such is the function of every final finish operation. Ensuring that final finish tire testing machines read and depict results correctly is a key step to assure the customer that they will be placing a quality tire with a desirable ride on their vehicle.

For more information about the types of Measurement Systems Analysis (MSA) tools and techniques mentioned in this paper, please contact Dr. Shaun M. Immel from Micro-Poise Measurement Systems, LLC. at simmel@micropoise.com.

Bibliography

Automotive Industry Action Group, Automotive Division. (2002). *Measurement Systems Analysis, Third Edition.* Milwaukee, WI: American Society for Quality.

Minitab Company Information. (2010). Retrieved July 27, 2010, from Minitab - Software for Quality Improvement. : http://www.minitab.com/en-US/company/default.aspx

Minitab Inc. (2010). *Assessing Measurement System Variation - Gage Studies for Continuous Data - Rel. 16, Ver. 1.0.* Retrieved August 2010, from Minitab, Inc.: http://www.mintab.com/uploadedFiles/Shared_Resources/Documents/Tutorials/TrainingSampleMeasurementSystems.pdf

Wheeler, D. J. (2009, January). *An Honest Gauge R&R Study*. Retrieved August 2010, from SPC Ink: http://www.spcpress.com