THE IMPACT OF MEASUREMENT SYSTEMS PERFORMANCE ON PROCESS CAPABILITY EVALUATION

Andrés Carrión
Universitat Politécnica de Valencia, Valencia, Spain
Angela Grisales
Centre for Quality and Change Management, Pereira, Colombia

Summary: *The evaluation of quality of products and processes depends on the ability to obtain accurate measurements consistently over time. With these measurements different control tools are used to monitor the performance of processes and the conformity of products. One of these tools, widely used and each time more recognized as relevant, is Capability analysis. Through a set of well-known parameters, as C_p and C_{pk}, capability is a key element in quality control with a continuous improvement approach. As capability indexes are computed using data coming from measurement processes, the consistency of capability indexes depends highly in the quality and reliability of the measurements obtained from such processes. In this paper we analyze the relation between the performance of the measurement systems and the representativity of the capability studies.*

Key words*: Measurement system analysis, capability, quality control, repeatability, reproducibility.*

1. INTRODUCTION

In manufacturing industry, the importance of measurement systems has been increasingly recognized. Data obtained using different types of gages and measurement devices (measurements) are the basis for all further quality analysis, and consequently the basis for a decision-making process with important economic and commercial consequences. Thus, all efforts must be done to guarantee the quality of measurements and the good performance of measurements systems.

A first step to achieve reliable measurements systems is to ensure that those systems have metrological traceability, usually achieved by the requirement of maintaining measurement devices in a state of calibration and maintain this calibration via periodic revisions. According to the International Vocabulary of Metrology [BIPM, 2012], calibration is the "operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication". These conditions usually are those of a metrological laboratory, and the people involved in this process are specialists in metrology.

But a part of this fact, that serves to ensure the performance of the instrument itself, it is also necessary to verify the actual performance of these devices when used in production conditions, by their actual users (production staff, not specialized in metrology).

Statistical Process Control (SPC) include two basic tools [Montgomery, 2008]. The first are the well-known control charts, centering the effort of achieving stable processes. Starting with the classic Shewhart charts, lot of different charts have been developed to improve the effectiveness of the control or to adapt to specific circumstances. The second basic tool is capability analysis, oriented no to process stability (considered here a pre-requisite) but to product/process conformity. That is, capability analysis evaluates the ability of the process to fulfill the requirements or specifications usually affecting the product (but sometimes the process itself). SPC tools use as input information coming from measurement processes, and in consequence the quality of these data affects the value of the analysis made with control charts, with capability analysis or whatever other tool used.

In this paper we will review the impact that the quality of data (or the performance of the measurement system) has over the value and representativeness of capability studies.

2. MEASUREMENT SYSTEMS EVALUATION

One the most used approaches to evaluate the performance of a measurement system comes from the automotive sector and is a quality core tool known as Measurements System Analysis [MSA, 2010]. The reference guide for using this tool has been developed by the Automotive Industry Action Group (AIAG), an association of carmakers and suppliers that include the three great North American OEMs: Ford, Chrysler and General Motors. There are other approaches as that generated by the German carmakers associated in the Verband der Automobilindustrie (VDA). In this paper we will use the AIAG scheme.

First it is important to understand what is a measurement system and what is the purpose of MSA. A measurement system is the combination of a measurement device, used by some specific users (or appraisers), in a specific way to measure one or more product characteristics under some conditions. In other words, is the set formed by *appraisers* (persons), measurement *instrument* and *method*, working in some environment. As not always environment can be controlled, usually the first three elements are considered the components of the measurement system: people, instrument and method. If environment were controlled, it can be included in method.

MSA defines the types of variations affecting a measurement system, thus reducing the quality of the data generated by measuring process (MSA 2010). These are:

- Affecting the value of the readings
  - Bias
  - Stability
  - Linearity
- Affecting the variation of the readings
  - Repeatability
  - Reproducibility
  - Consistency
  - Uniformity
- Affecting the sensitivity of the readings
  - Number of distinct categories (ndc)

All types of variation are important, and a good, in deep analysis of a measurement system requires to evaluate all of them. Only the number of distinct categories, ndc, can be discussed, as it has an exact relation with % GRR, and in consequence may be ignored [Carrion, 2013]. Unfortunately, by different reasons (time availability, lack of knowledge, budget limitations, short term thinking …) decisions tend to be taken considering only one or two of them. In any case the more relevant is the combination of repeatability and reproducibility in the R&R or GRR index.

Repeatability is usually interpreted as the variability in measurements caused by the instrument and is referred to as EV (equipment variation) or σ_e (standard deviation caused by the equipment). Statistically repeatability is the standard deviation generated when one appraiser, with one instrument measures repeatedly one part.

The combination of repeatability and reproducibility is the variation in measurements due to the measurements system, and is known as R&R or GRR, as mentioned above. The expression is:

$$GRR^2 = EV^2 + AV^2 \tag{1}$$

Reproducibility is interpreted as the variation due to the fact that more than one appraiser is taking part in the measurement process, or simply the variation caused by appraisers. It is noted as AV (appraiser variation) or σ_a (standard deviation due to the differences between appraisers). Statistically is the standard deviation between appraisers' means, corrected with the effect of repeatability. Two other variabilities have to be considered to complete the analysis of the measurement system: the total variation TV, or σ_t, the standard deviation observed in the measurements obtained; and the part variation PV or σ_p, the standard deviations between parts.

Usually GRR analysis is performed using a experimental design that includes three appraisers, ten parts and three repeated readings (per appraiser per part). The better way for analyze resulting data is ANOVA. In terms of this type of analysis we have two controlled factors (appraiser and part), one with three levels (the three appraisers) and the other one with ten levels (the ten parts to be measured) and with three repetitions, being the response variable the reading obtained. The general ANOVA table do not generates directly the values of EV, AV, PV and TV, but simple well-known expressions can be used. table 1 shows a general ANOVA table for two factors without considering interaction, and table 2 with interaction.

Table 1. ANOVA table without interaction

Source	Sum of squares	Degrees of freedom	Mean square	F-Ratio	P-value
Appraisers	SSa	a-1	MSa	MSa/MSr	
Parts	SSp	p-1	MSp	MSp/MSr	
Residual	SSp	a*p*r-p-a+1	MSr		
Total	SSt	a*p*r-1			

Table 2. ANOVA table with interaction

Source	Sum of squares	Degrees of freedom	Mean square	F-Ratio	P-value
Appraisers	SSa	a-1	MSa	MSa/MSr	
Parts	SSp	p-1	MSp	MSp/MSr	
Interaction	SSi	(a-1)*(p-1)	MSi	MSi/MSr	
Residual	SSp	a*p*(r-1)	MSr		
Total	SSt	a*p*r-1			

If interaction between appraisers and parts is present, the variability of the measurements system must include this term, resulting that equation 1 should be:

$$GRR^2 = EV^2 + AV^2 + I^2 \qquad (2)$$

The most important relation for the analysis of the variability of a measurement system express the relationship between all four variabilities mentioned:

$$TV^2 = PV^2 + GRR^2 \qquad (3)$$

Where GRR is obtained using equation 1 or eq. 2, depending on the presence of interaction.

Alternatively, we can write this expression using variances (is just a matter of notation, as TV, PV or GRR are standard deviations):

$$\sigma_t^2 = \sigma_p^2 + \sigma_m^2 \tag{4}$$

Where, σ_m comes from rewriting equation 1:

$$GRR^2 = EV^2 + AV^2 \rightarrow \sigma_m^2 = \sigma_e^2 + \sigma_a^2$$

Or from equation 2:

$$GRR^2 = EV^2 + AV^2 + I^2 \rightarrow \sigma_m^2 = \sigma_e^2 + \sigma_a^2 + \sigma_i^2$$

Have in mind that in both cases the result that is of interest for this paper is eq. 4, and do not changes.

The acceptance criterion for a measurement system, based on the variability assessment is in fact based in the percentage of TV covered by the measurements system error, GRR:

$$\%GRR = \frac{GRR}{TV} * 100 \tag{5}$$

A measurement system is generally accepted if %GRR is under 10%, can be accepted in some cases if it is under 30%, and cannot be accepted if is over 30%.

3. PROCESS CAPABILITY

Process capability studies seek to assess the ability of the process to meet specifications. They basically make sense in measurable quality characteristics (length, dimensions, temperature, pressure, among others).

Before determining the Capability of the process, it must be ensured that the process is under statistical control, that is, it is a stable process. To ensure the above, special causes of variation must be eliminated. These special causes generate changes in the variation, location, and shape of the processes. The data obtained should be used to predict the future performance of the process [Montgomery, 2008].

3.1 Normal Data Capability

The C_p index indicates potential process capability. It compares the maximum allowable variation as indicated by the tolerance to the width of the process distribution (i.e., process variation).

The C_p provides measure of the extent to which a process will produce output which meets specifications if the production output distribution is centered between the specification limits. C_p is not impacted by process location (i.e. centering) and can only be calculated for bilateral tolerances. It has no meaning for unilateral tolerances.

$$C_p = \frac{USL - LSL}{6 * \sigma} \qquad (6)$$

where:

C_p = Potential process capability
USL = Upper specification limit
LSL = Lower specification limit
σ = Standard deviation measurements

C_p indicates how many times the width of process sample distribution can fit between the specification limits. As the process variability decreases the C_p index increases. Therefore, larger C_p values are better than smaller C_p values. Some frequently considered threshold values are 1.33 and 1.67. Usually the first one is the minimum required C_p and the second one is the preferred minimum value (see table 3). Same critical values will apply for C_{pk}.

Table 3. Threshold values for C_p

Potential Process Capability	C_p Value
Bad	< 1.33
OK	1.33 to 1.66
Good	>1.66

3.2 C_{pk} Index

The C_{pk} index indicates actual process capability. It takes the process location as well as the process variation into account. It used to determine whether or not a process is capable of meeting customer requirements. It also uses C_{pk} with C_p to predict potential process capability.

To calculate C_{pk} for bilateral tolerances as the minimum of the lower capability index (denoted CPL or $C_{pk(LSL)}$) or the upper capability index (denoted CPU or $C_{pk(USL)}$) as follows:

C_{pk} = Minimum CPL, CPU,

$$C_{pk} = \text{Minimum} \left[CPL = \frac{\bar{x} - LSL}{3 * \sigma},\ CPU = \frac{USL - \bar{x}}{3 * \sigma} \right]$$

where:

$\bar{x}$ = Process average
LSL = Lower specification limit
USL = Upper specification limit
σ = Standard deviation measurements

For unilateral tolerances, C_{pk} is equal to the lower capability index or the upper capability index, depending on whether the tolerance is an LSL or USL. Therefore, it calculates C_{pk} for a unilateral tolerance using one or other formula, CPL or CPU.

3.3 Interpreting C_p and C_{pk}

Using C_p and C_{pk} together provides information about the process centering:

- When C_p is equal to or approximately equal to C_{pk}, the process is well centered on the target.
- A large discrepancy between C_p and C_{pk} indicates a process centering problem. If the process is off target (i.e., $C_{pk} < C_p$), it can make adjustments to achieve the "potential" indicated by C_p (i.e., $C_{pk} = C_p$).

4. IMPACT OF GRR OVER CAPABILITY INDEXES

Capability indexes are computed with data coming from measurement processes, and in consequence are affected by the quality of these data. But it is of interest to formalize this relationship, and this is the aim of this paper.
In all capability indexes, we find the standard deviation of the process denoted as σ. This standard deviation is calculated from the measurements and corresponds in MSA terminology to the total variation, TV or σ_t. Then we can write, from Eq. 6:

$$C_p = \frac{USL - LSL}{6 * TV}$$

And, according to eq. 4:

$$C_p = \frac{USL - LSL}{6 * \sqrt{PV^2 + GRR^2}}$$

Where we find a first insight of the impact of GRR over C_p, and similarly over C_{pk}.
It can be developed that:

$$\frac{1}{C_p^2} = \frac{36 * (PV^2 + GRR^2)}{(USL - LSL)^2} = \frac{36 * PV^2}{(USL - LSL)^2} + \frac{36 * GRR^2}{(USL - LSL)^2}$$

And, we can name:

$$\frac{1}{C_{pP}^2} = \frac{36 * PV^2}{(USL - LSL)^2} \qquad \frac{1}{C_{pSM}^2} = \frac{36 * GRR^2}{(USL - LSL)^2}$$

Thus:

$$\frac{1}{C_p^2} = \frac{1}{C_{pP}^2} + \frac{1}{C_{pSM}^2} \qquad (7)$$

Or:

$$C_p = \sqrt{\frac{1}{\frac{1}{C_{pP}^2} + \frac{1}{C_{pSM}^2}}}$$

The direct capability index C_p can be considered as the *Observed Capability*, while C_{pP} is the *Actual Capability*, that is the capability obtained if we had a perfect (free of error) measurement system. The third capability involved in Eq. 7 is the *Measurement System Capability*, and is linked with the gage repeatability and reproducibility, both in absolute value, GRR, or in percentage of the Total Variation, % GRR.

The practical meaning of this relationship is that always, the observed capability will be lower than the actual capability, due to the error introduced by the measurement system. The size of this difference depends on the performance of the measurement system.

Using Eq. 7, we can build table 4, that presents the C_p values for different combinations of C_{pP} and C_{pSM}. Special emphasis has been made in values of C_{pP} usually considered as critical: 1.33 and 1.67.

Table 4. C_p values after C_{pP} and C_{pSM}

		C_{pP}								
		0,5	1	1,33	1,67	2	3	4	5	6
C_{pSM}	0,5	0,35	0,45	0,47	0,48	0,49	0,49	0,50	0,50	0,50
	1	0,45	0,71	0,80	0,86	0,89	0,95	0,97	0,98	0,99
	2	0,49	0,89	1,11	1,28	1,41	1,66	1,79	1,86	1,90
	4	0,50	0,97	1,26	1,54	1,79	2,40	2,83	3,12	3,33
	6	0,50	0,99	1,30	1,61	1,90	2,68	3,33	3,84	4,24
	7	0,50	0,99	1,31	1,62	1,92	2,76	3,47	4,07	4,56
	10	0,50	1,00	1,32	1,64	1,96	2,87	3,71	4,47	5,14
	13	0,50	1,00	1,32	1,65	1,98	2,92	3,82	4,67	5,45
	16	0,50	1,00	1,33	1,66	1,98	2,95	3,88	4,77	5,62
	19	0,50	1,00	1,33	1,66	1,99	2,96	3,91	4,84	5,72
	20	0,50	1,00	1,33	1,66	1,99	2,97	3,92	4,85	5,75
	25	0,50	1,00	1,33	1,66	1,99	2,98	3,95	4,90	5,83
	30	0,50	1,00	1,33	1,66	2,00	2,99	3,96	4,93	5,88

We can also obtain the expression that directly links observed and actual capabilities, depending on the % GRR value:

$$C_p = \frac{\%GRR}{100} * C_{pSM} \qquad (8)$$

Using Eq. 7 and Eq. 8, the direct relationship between C_p and C_{pP} can be obtained:

$$C_p = C_{pP}\sqrt{1-\left(\frac{\%GRR}{100}\right)^2} \qquad (9)$$

This relation produces the graphic in figure 1, were % GRR=0 corresponds to the behavior of a "free of error" measurement system, where observed and actual capabilities will be the same. As it can be seen, for % GRR under 30 the impact in observed capability is very low.

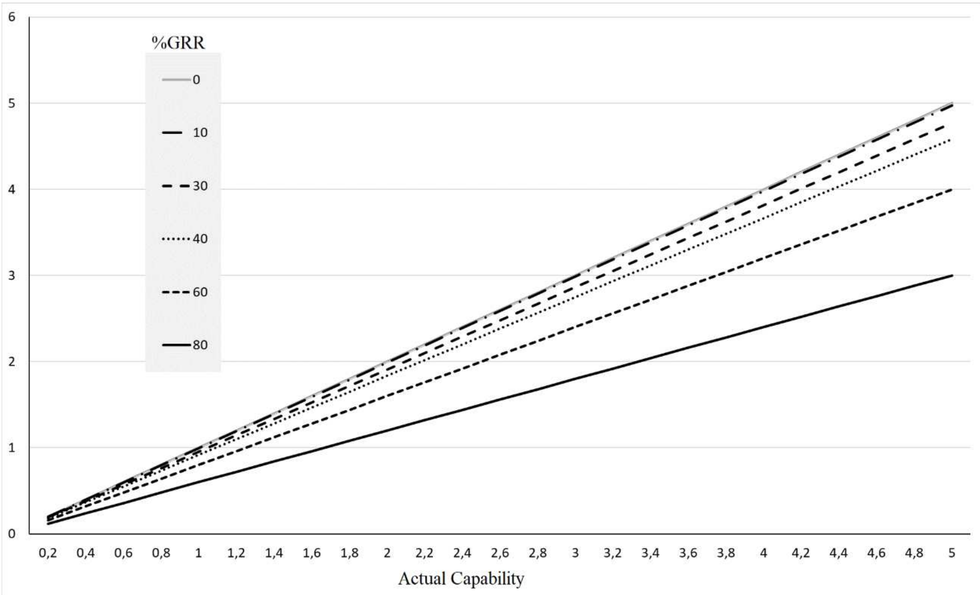

Figure 1. Observed Capability versus Actual Capability, for different %GRR

For high % GRR values, the impact of measurement system lack of performance over the capability evaluation may be relevant. For instance, with a % GRR of 40, an actual capability of 2,5 is reduced to 2, or with a %GRR of 80, an acceptable actual capability of 1.66 is reduced to a non-acceptable value of 1.

The magnitude of this deviations is presented in table 5, where percentage errors between actual and observed capabilities have been computed.

These percentage errors between actual and observed capabilities can be linked with the values of % GRR. These errors, E%, presented in table 2, are obtained with a simple expression:

$$E\% = \frac{C_{pP} - C_p}{C_{pP}} * 100$$

Table 5. Percentage of deviation between C_p and C_{pP}

		C_{pP}								
		0,5	1	1,33	1,67	2	3	4	5	6
C_{pSM}	0,5	29,3	55,3	64,9	71,3	75,7	83,6	87,6	90,0	91,7
	1	10,6	29,3	40,0	48,6	55,3	68,4	75,7	80,4	83,6
	2	3,0	10,6	16,8	23,2	29,3	44,5	55,3	62,9	68,4
	4	0,8	3,0	5,1	7,7	10,6	20,0	29,3	37,5	44,5
	6	0,3	1,4	2,4	3,6	5,1	10,6	16,8	23,2	29,3
	7	0,3	1,0	1,8	2,7	3,8	8,1	13,2	18,6	24,1
	10	0,1	0,5	0,9	1,4	1,9	4,2	7,2	10,6	14,3
	13	0,1	0,3	0,5	0,8	1,2	2,6	4,4	6,7	9,2
	16	0,0	0,2	0,3	0,5	0,8	1,7	3,0	4,6	6,4
	19	0,0	0,1	0,2	0,4	0,5	1,2	2,1	3,3	4,6
	20	0,0	0,1	0,2	0,3	0,5	1,1	1,9	3,0	4,2
	25	0,0	0,1	0,1	0,2	0,3	0,7	1,3	1,9	2,8
	30	0,0	0,1	0,1	0,2	0,2	0,5	0,9	1,4	1,9

Substituting here C_p by its expression according to Eq. 9, we obtain:

$$E\% = \left(1 - \sqrt{1 - \left(\frac{\%GRR}{100}\right)^2}\right) * 100$$

Thus, the percentage error between observed and actual capabilities depends only on % GRR value. If we look at the critical values of % GRR, 10 to consider the measurement system as acceptable and 30 for a conditional acceptance, we find that:

- When % GRR is equal or lower than 10, the percentage error E% is lower than 0.5%.
- When % GRR is lower than 30 the percentage error E% will be lower than 4.6%.

5. CONCLUSIONS

In this paper we haver reviewed the impact of the lack of performance of measurement systems over capability evaluation. We have used as an indicator of the performance of the measurement system the gage repeatability and reproducibility GRR and its value as a percentage of the total process variation, % GRR. To evaluate capability, we have used the well-known Cp, potential capability index.

The exact expressions linking potential capability index Cp with GRR and % GRR have been developed, allowing a detailed analysis of the relationship. We have shown that if % GRR do not accomplishes the threshold value of 30, relevant impacts may affect the capability evaluation, arriving to deviations from the actual capability of 80% and 90%.

Also, we have demonstrated that critical values for % GRR correspond to specific values for the percentage error in the calculation of C_p (observed capability) with respect to C_{pP} (actual capability). When % GRR is under 30, the percentage error in calculation of Cp is lower than 4.6%; when % GRR is under 10, the percentage error for C_p is lower than 0.5%. Simplifying, and as a rule of thumb, we can remember that 30% GRR means less than 5% estimation error in C_p and 10% GRR means less than 0.5%.

As final comment, we can remark that when % GRR is in acceptable range (0 to 10%, and 10 to 30%), the impact of the measurement system performance over the capability estimation is not relevant. But if % GRR is over 30%, impact can be very important, invalidating the capability estimations and thus showing and reinforcing the value of having quality measurement systems, that produce reliable, accurate and significant data.

REFERENCES

[1] AIAG (Automotive Industry Action Group), 2010. AIAG reference manual, Measurement systems analysis. MSA. Fourth ed., AIAG, Southfield, MI.

[2] BIMP (Bureau International des Poids et Mesures), 2012. International vocabulary of metrology - Basic and general concepts and associated terms (VIM). 3rd edition. JCGM 200:2012. BIMP, Paris.

[3] Carrión, A., Grisales del Río, A., 2013. Number of distinct data categories and gage repeatability and reproducibility. A double (but single) requirement. Measurement, Volume 46, Issue 8, October 2013, pp. 2514-2518.

[4] Montgomery, Douglas C., 2008. Introduction to Statistical Quality Control. John Wiley and Sons.