

Methods of Measurement System Quality Assessment in Case of Two Devices

Krzysztof Kowal and Michał Szymczak*

Department of Statistics, University of Łódź, The Faculty of Economics and Sociology, 90-255 Łódź, 3/5 POW Street, Poland

Abstract: The article discussed the selected methods of measurement systems analysis (MSA) in case of different devices usage. Analysis of variance is a common method, widely accepted and applied in industry to analyze measurement systems by taking into consideration different sources of variability: equipment, operators, parts and their interaction. Regression is a method which could simplify the analysis by shortening its time and decreasing sample size. It also may enable more clear and suitable answer. The included case study concerning two coordinate measuring machines (CMM) indicates the usefulness of the regression as a method for high precision and automated measurement systems comparison. Some actions resulting from the inferences can be undertaken by managers and engineers in industrial enterprises to reduce the cost of double measurements.

Keywords: ANOVA R&R, linear regression, measurement system analysis (MSA), SPC.

INTRODUCTION

Since statistics started to be appreciated also in industrial engineering, some of the common statistical methods are applied in the new environment. Several of the procedures were implemented with changed names, but some of them needed to be modified to be valuable. Especially, it happens in different conditions with certain assumptions appropriate for measurements in production. Measurements are conducted by engineers for many reasons, such as control of production processes, quality control of products and components, designing new processes and appliances or others. One of the most interesting issue apart from the inferences from the observed process is the measurement error [1].

Control charts are basic methods applied to production processes supervision. Many authors like Chambers [2], Kończak [3], Wheeler [4] pay special attention to correct reaction on signals from the charts which may save a process from becoming unstable.

One of the most dangerous actions done sometimes by engineers is too fast reaction on signals from a control chart without knowing the possible measurement error. This may lead to a catastrophe. Hence, some methods of measurement system quality assessment were proposed. Considering a measurements of a process we can divide its total observed variation into two main sources:

$$\sigma_{total}^2 = \sigma_{process}^2 + \sigma_{measurement\ system}^2, \quad (1)$$

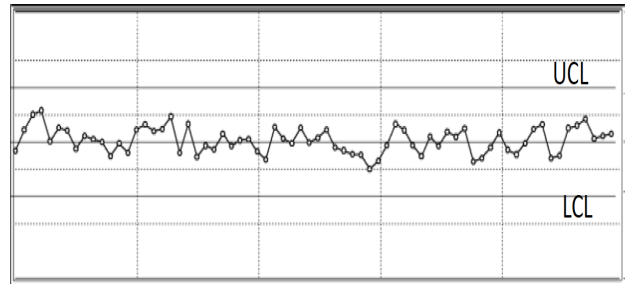


Figure 1: Control chart for individual values of an in-control process - view in SNAP system.

Source: Own elaboration.

where $\sigma_{process}^2$ is the true variation of produced parts, and $\sigma_{measurement\ system}^2$ is the variation of the measurement system. A good measurement system is the one which has the less influence on the total variation. Thus, we are interested in minimization of

$$\frac{\sigma_{measurement\ system}^2}{\sigma_{total}^2} \quad (2)$$

We can classify measurements variability into different categories (Figure 2).

The problem analyzed in the paper regards a common situation when the same process is measured and monitored by two different measurement systems. Equipments applied in both systems can be similar, but whole systems may still differ significantly. The consequence of the situation is the possibility of taking wrong decisions leading to change a predictable process into unpredictable and uncontrolled. This, in turn, may cause financial consequences arising from producing parts which are out of tolerances with their

*Address correspondence to this author at the Department of Statistics, University of Łódź, The Faculty of Economics and Sociology, 90-255 Łódź, 3/5 POW Street, Poland; Tel: +48 504 444 712; E-mail: micszym@gmail.com

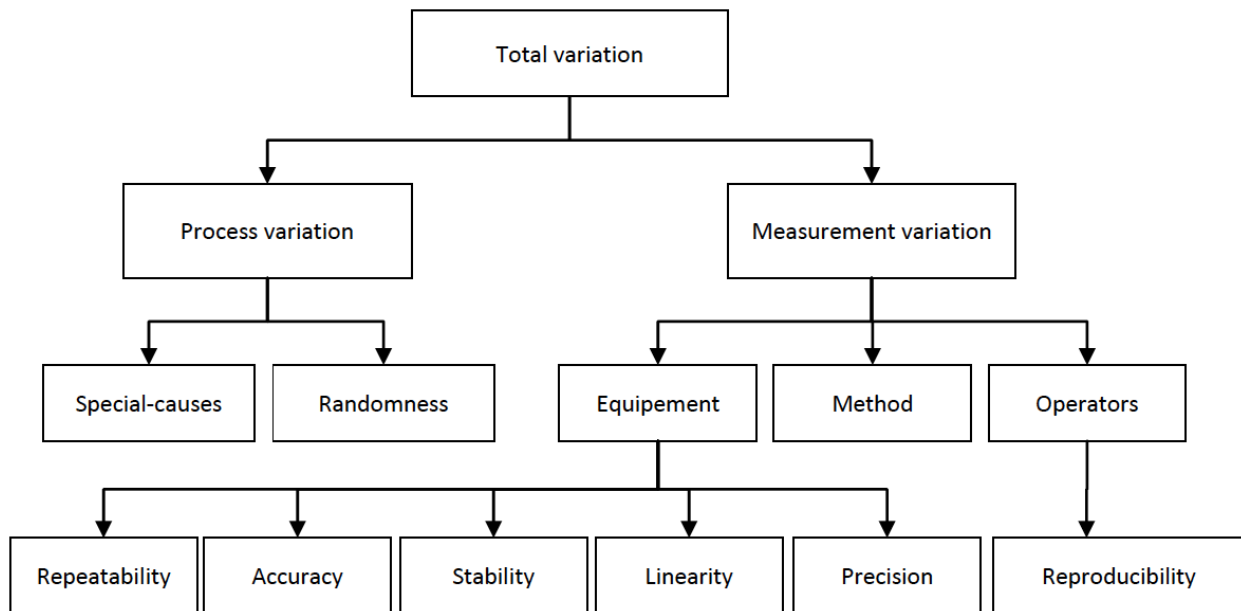


Figure 2: Different sources of variation observed in production process measurements.

Source: Own elaboration.

true dimensions and assembly problems. Moreover, if such a problem arises when comparing customer and supplier measurement systems there is a risk of unnecessary claims and conflicts. In the projects, which introduce new designs of parts or products, one of the essential issues is time. Double measurements resulting from lack of trust and different systems causes delays in releasing processes.

In reality, the problem is very complex. Managers and engineers usually tend to work based on standardized solutions published in internal, national or international norms such as ISO or similar. Measurement system analysis methods are normalized in the MSA norm. However, guidelines included in the norm does not specify exactly means of analysis for specific cases. Most types of approaches consider situations in which one device is used by many operators. For the case as raised in the paper, there are two systems (devices) applied for process control. This requires some modifications of measurement systems assessment techniques.

One of the critical results coming from discrepancies between different control systems are delays in new components releases. Reasons of the problem are connected with measurements which are contradictory. If the new product is going to be produced internally, the significance of such an issue is less. There still exist a possibility to agree that only one system will be consequently applied to monitor the process variation. Such management decision will be,

however difficult if the problem concerns cooperation with a supplier. Then releasing parts requires agreement that they are produced according to specifications. Practice shows many examples where even the most important dimensions when measured differently cannot be assessed the same. Extreme situations are when one of the systems shows results outside lower tolerance, and the second above the upper limit. It really causes interpretational troubles. Secondly, lengthen necessary time for the whole project very often means also special preparations such as building stocks in rented magazines, transportation etc. Summarizing all the things together, fast and simple method appropriate to implement in this environment should be welcomed. By improving comparability of the results from the systems, time reduction of the process should be possible. If the relationship between them is known, measurements conducted by one of the side can be reduced.

It is easily noticeable that total observed variance of a process consists of two main sources – variability of parts dimensions and measurements error. The second source is considered in the article in the context of using two different measurement systems to monitor the same process. Variability sources connected with measurement systems are defined here as follows:

Repeatability – differences in results obtained by an operator with the usage of a specific measurement system when measuring the same parts,

Accuracy – difference of the measurements average of a component from its true value (the biasness),

Stability – ability of a measurement system to maintain its properties in time,

Linearity – ability of a measurement system to keep the same error independently of what is the true value,

Precision – property of a measurement system related to deviations of results from the average result of a part,

Reproducibility – differences in results obtained by one measurement system compared with results of another system when measuring the same parts.

The influence of several factors such as biasness, stability, linearity and precision on measurement errors is normally controlled by systematical machines calibrations. Some of the problems, however cannot be simply solved with this movement. This regards the differences in results when using devices of the same quality. The problem then is to chose the fastest and the cheapest way of assessing quality of measurements which fulfills requirements of taking correct decision according to certain criteria. Hence, a simple criterion is proposed to distinguish between acceptable and unacceptable measurement systems when using regression method and is compared to classical approach of measurement system analysis described widely in Automotive Industry Action Group (AIAG) [5], and Dietrich, Schulze [6].

Highly precise and automated Coordinate Measuring Machines (CMM) are used for measuring parts in the analyzed case study. The meaningful information is that both systems are quite expensive (costs of a new equipment starts at tens of thousands USD). The measurement errors certified by producers are given in tens of microns¹. Still, however very different results are possible. The easiest solution would be to treat one of the system as always better – a master system according to which the second one is adjusted and calibrated. Here, however the quality of both can be similar and high. They may differ according to their construction details, producers, software, touch triggers, programs etc. Thus, in such situation it is necessary to manage with two equivalent systems, where none of them can be considered as better.

1. MEASUREMENT SYSTEM ANALYSIS USING ANOVA FOR TWO TREATMENT DESIGN

Analysis of variance is a method recommended for measurement system analysis as the most precise and valuable, compared for example to an average and ranges method [7]. The arguments are because of the interaction factor between parts and operators which is also included in a research. This is a kind of standard analysis of measurement systems according to Automotive Industry Action Group (AIAG), which is applied also in many other industries.

In order to understand why this approach is adopted as a standard, it is necessary to know the realities of a common situation, in which it is used. There are two measurement conditions considered in the analysis. The first is typically the influence of appraisers, mostly related to reproducibility and the second one is repetition (repeatability) related to an equipment quality. In such case, the ANOVA Table is built for two factors. When analyzing a case with more than one equipment we would need three factor design ANOVA. The case is restricted to the usage of automatic devices, which minimize the influence of human factor. Such an example requires some modification of ANOVA calculations and interpretation. If highly automated systems are applied, by definition the influence of operator is very little. Assuming this, the factor related with an appraiser can be replaced by an information about the measuring system, because this is then a measurement condition which changes in the system. In such case, the ANOVA Table is built as follows (Table 1).

Where $i = 1, \dots, n$ is the number of parts, $j = 1, \dots, k$ is the number of devices (equipment) and $r = 1, \dots, m$ is a number of measurement repetitions. In order to assess the quality of the whole system it is needed to estimate and interpret the following indicators.

Such an approach requires at least two repetitions and two devices to enable estimation of both factors influence. Thus, the total cost of research grows when increasing number of devices used and repetitions. The first column names the sources of variation. Equipment variation (EV) can be connected with the possibility to obtain results close to the previous. The higher value is observed, the worst is the system itself. SV (measurement systems variation) is the result of variability caused by differences in average results obtained by different devices. The higher estimate – the poorer is the comparability between analyzed systems. The interaction term (INT) is the estimator of the interaction between values obtained for

¹Maximum error permitted by producers according to EN ISO 10360-2:2002 is given in micrometers.

Table 1: Anova Table for Gage Repeatability and Reproducibility Source

	Degrees of freedom	Sum of squares	Mean square	F
Device	$k - 1$	$SS_D = n \cdot r \cdot \sum_{j=1}^k (\bar{x}_{j\cdot} - \bar{x}_{\cdot\cdot})^2$	$MS_D = \frac{SS_D}{(k-1)}$	$\frac{MS_D}{MS_{DP}}$
Part	$n - 1$	$SS_P = k \cdot r \cdot \sum_{i=1}^n (\bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot\cdot})^2$	$MS_P = \frac{SS_P}{(n-1)}$	$\frac{MS_P}{MS_{DP}}$
Interaction	$(n-1)(k-1)$	$SS_{DP} = TSS - (SS_P + SS_D + SS_R)$	$MS_{DP} = \frac{SS_{DP}}{(n-1)(k-1)}$	$\frac{MS_{DP}}{MS_R}$
Repeatability	$nk(r-1)$	$SS_R = \sum_{i=1}^n \sum_{j=1}^k \sum_{m=1}^r (x_{ijm} - \bar{x}_{ij\cdot})^2$	$MS_R = \frac{SS_R}{nk(r-1)}$	
Total	$nkr - 1$	$TSS = \sum_{i=1}^n \sum_{j=1}^k \sum_{m=1}^r (x_{ijm} - \bar{x}_{\cdot\cdot\cdot})^2$		

Table 2: Variance Estimators

Source of variation	Estimator
EV – Equipment Variation (repeatability)	$\tau^2 = MS_R$
INT - Interaction	$\gamma^2 = \frac{MS_{DP} - MS_R}{r}$
SV – Measurement System Variation (reproducibility)	$\omega^2 = \frac{MS_D - MS_{DP}}{nr}$
PV - Part Variation	$\nu^2 = \frac{MS_P - MS_{DP}}{kr}$

Source: Own elaboration based on Automotive Industry Action Group (AIAG), 2010. Measurement Systems Analysis Reference Manual, 4th edition. Chrysler, Ford, General Motors Supplier Quality Requirements Task Force.

components and applied device. In order to assess a system, first $GRR = \tau^2 + \gamma^2 + \omega^2$ indicator is calculated and then estimator of $\%R \& R = \frac{GRR}{GRR + PV}$ is taken into consideration for final assessment.

1.1. A Case Study with the Usage of Two Coordinate Measuring Machines

Ten parts were randomly chosen from the production process. Each of them was measured twice using two different machines operated by a metrologist in a

laboratory. Summarizing, the total number of measurements was equal to 40. Presented example concerns a situation, where two different machines are used for process control, assessment of product quality and if it meets the specification. The product is a metal part, with its length as one of the most important characteristic (critical to quality) for assembly and a customer. Here the influence of an operator is eliminated by automatization, so this factor is replaced by an information about a device. ANOVA results were obtained with the usage of Minitab program [8]. Calculations for the example are presented in Table 3.

Table 3: ANOVA Results for a System Analysis with the Usage of Two Measuring Devices

Two-Way ANOVA Table With Interaction					
Source	DF	SS	MS	F	P
Part	9	0,263262	0,0292513	82,389	0,000
Machine	1	0,006600	0,0065996	18,588	0,002
Interaction	9	0,003195	0,0003550	163,727	0,000
Repeatability	20	0,000043	0,0000022		
Total	39	0,273100			

Gage R&R		
Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0,0004908	6,36
Repeatability	0,0000022	0,03
Reproducibility	0,0004887	6,33
Machine	0,0003122	4,05
Interaction	0,0001764	2,29
Part-To-Part	0,0072241	93,64
Total Variation	0,0077149	100,00

Source	StdDev (SD)	Study Var (6 * SD)	%Study Var (%SV)
Total Gage R&R	0,0221547	0,132928	25,22
Repeatability	0,0014726	0,008835	1,68
Reproducibility	0,0221057	0,132634	25,17
Machine	0,0176700	0,106020	20,12
Interaction	0,0132829	0,079697	15,12
Part-To-Part	0,0849945	0,509967	96,77
Total Variation	0,0890525	0,534315	100,00

Source: Own elaboration.

Contrary to the standard analysis of measurement system where different operators are considered, here there are two different machines. Thus, the interpretation of reproducibility is a difference of the results obtained with different machines.

To interpret the results in the context of a straightforward decision about the system, it is necessary to compare them with certain criteria. According to AIAG guidelines, the limit of total measurements error acceptance when studying variances is maximum 9% if the gage R&R (repeatability and reproducibility) variation (τ^2) is divided by total variation ($\tau^2 + v^2$), however 1% is recommended. Because of the interpretation utilities, it is common to compare standard deviations. Then the corresponding criterion is 30% with the recom-

mendation of 10% for. $\frac{\sqrt{\tau^2}}{\sqrt{\tau^2 + v^2}}$ In the analyzed case,

the highest influence on the measurement variation has reproducibility. It means, that the most important factor in the model is a device used for measurements. Summarizing, although the results show that system is conditionally acceptable (6,36% and 25,22% for variations and standard deviations analysis indicators), in fact the analysis does not give the exhaustive

information about what is really important in the whole system – statistical relationship between measurements from two different machines.

In case of serial production, measurements often has not the highest priority. One of the biggest challenge for a researcher is to collect data. That is why sample sizes are frequently reduced to minimum. It also should be noticed that a measurement costs time and money. All these factors in consequence make an atmosphere where some optimizations are strongly recommended.

2. MEASUREMENT SYSTEM ANALYSIS WITH THE USAGE OF LINEAR REGRESSION

When analyzing issues related with real processes in a production enterprise, it is impossible to ignore very specific atmosphere accompanying a research. There is continuously time pressure on results and willingness of most costs reduction.

Linear regression is a well-known method commonly used in modeling relationship between variables. Considering two variables, the model reduces to a special case with one regressor and a regressant. The question which is put when analyzing

Table 4: Regression Equation Estimating Linear Relationship between Two Measurement Systems

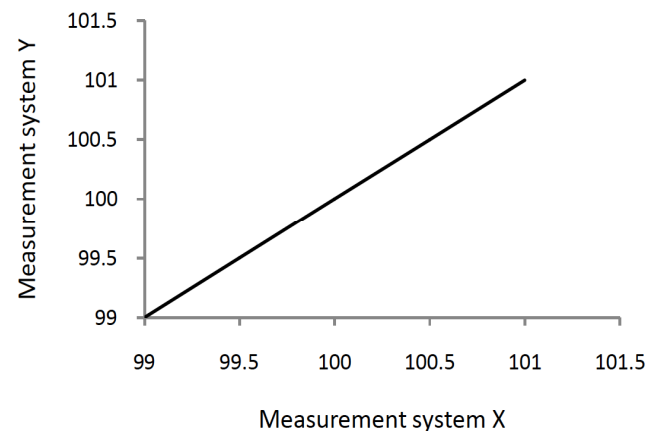
Regression Equation						
$Y = -2,40889 + 1,03046 X$						
Coefficients						
Term	Coef	SE Coef	T	P	95% CI	
Constant	-2,40889	4,56816	-0,5273	0,612	(-12,9431; 8,12531)	
X	1,03046	0,05823	17,6975	0,000	(0,8962; 1,16473)	
Summary of Model						
SE = 0,0151730		R-Sq = 97,51%		R-Sq(adj) = 97,20%		

two measurement systems sounds “What result will we get if we use a measurement system on condition that we have a certain result from another one?”. Similar approach was discussed by several authors [9-13]. The question leads to the analysis of results relationship between different means of measurement. Estimation of the function $y = f(x)$ requires an assumption of which variable is dependent and independent. However, both systems are completely independent. Thus, a decision of which variable is to be placed on which side of the model depends only on the researcher’s choice. A logical approach, however would recommend placing on the left hand side the variable describing result from a system treated as superior (e.g. customer’s system or a system that historically was used to assess the quality of a measured part as the first or more often). Then the estimates of parameters characterizing the influence of a regressor can be seen as an assessment of how good is the quality of the new system to reflect the current system results. Although, the only thing which should determine measurement result should be the true value, statistical relationship should be significant.

The model $y_i = a + bx_i + \xi_i, i = 1, \dots, n$ describes a linear relationship between two systems and ordinary least squares estimates gives the approximation of the function. If the parameters meets the requirements $a=0$ and $b=1$, then such a relationship can be considered as the ideal state.

A well-known t-test can be applied to test the relationship. If there is no ground to reject both hypotheses of $a=0$ and $b=1$ with the assumed level of significance, then the measurement systems are assumed to be equivalent. For the considered

example, the results based on 10 measurements from each machine obtained with the usage of Minitab program are as follows (Table 4).

**Figure 3:** Ideal linear relationship between two measurement systems.

Source: Own elaboration.

Hence, we can easily see that in the case of the example, the t-test for the constant term implies that with the level of significance $\alpha = 0,05$ there is no ground to reject the hypothesis that $a = 0$. The same t-test for $b = 1$ results with evaluated t statistic $t = 0,523$. Compared to the critical value of t^* with 8 degrees of freedom with the same level of significance, we have $t^* = 2,3060 > 0,523$, what implies that there is no ground to reject the null hypothesis. Also estimates of R^2 and R_{adj}^2 indicate that the second measurement variation is very well explained by the first measurement variation. If the reproducibility of the systems is proper, then this inference is quite obvious since both variations should be equal. Hence, from statistical point of view instead of evaluating R^2 the

better choice might be to test equality of the first and second measurements variances.

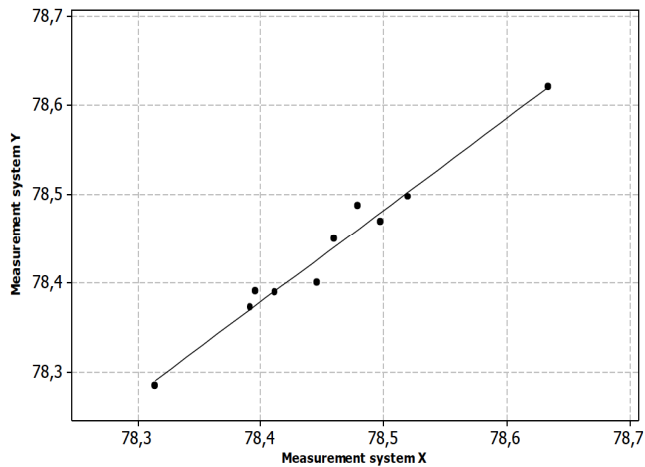


Figure 4: Relationship between measurement systems with linear regression.

In order to fully assess the quality of the measurements and answer the stated question, residuals from the model should be analyzed more carefully. The distribution of the residuals is characterized by Figure 5.

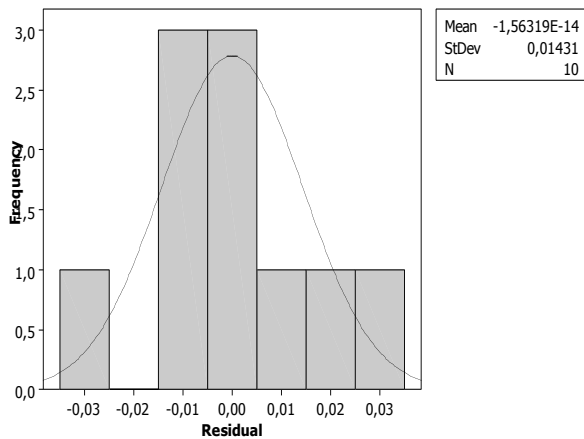


Figure 5: Distribution of residuals from the model.

From both, practical and theoretical point of view, the distribution of residuals is important. To verify a hypothesis of a normal distribution of residuals, a Shapiro-Wilk test was applied. W statistic was obtained ($W = 0,96$) and compared with the critical value of $W_{0,05} = 0,84$ resulting with the inference that there is no ground to reject the null hypothesis.

Although the measurement systems seem to be comparable, there are some differences in results because of existing sources of errors. Proposed criterion corresponding to the one from MSA norm, for measurement systems acceptance can be

$$\theta = \frac{SE}{\hat{\sigma}} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 (n-1)}{\sum_{i=1}^n (y_i - \bar{y})^2 (n-2)}, \tag{3}$$

where SE is the standard error from the model, and $\hat{\sigma}$ is the estimator of total standard deviation from measurements.

Such a defined indicator can very simply and exhaustively show the share of uncertainty coming from specific measurement error compared to the total variation. Different level of acceptance may be considered. If $\theta \leq 10\%$, the system could be fully acceptable, if $10\% < \theta \leq 30\%$, then the decision of acceptance would depend on the importance of the measured feature. In case of the example, the result of the criterion is

$$\hat{\theta} = \frac{0,015173}{0,089053} \approx 0,17 = 17\%,$$

meaning that the whole system consisted of two machines could be accepted.

3. SUMMARY OF THE RESEARCH ORGANIZATION IN PRACTICE

When comparing customer's and supplier's measurement system, sometimes is not defined how the critical trait should be measured. In practice drawings are frequently incomplete and it happens that such information are missing. Common method for both sides are necessary then to establish. The most appropriate time is during enabling supplier to produce components. Time pressure however causes hurry and finally, process is often released without standardizing the measuring method.

Improvements of measurement systems can be also done irrespectively of what happened in the past. To make it effectively, the first step is to define the problem and check if it really exist. Suppose that there are two devices applied to monitor a process. Possible ways and their effects are presented in Figure 6.

When having different systems, even if they are expensive and automatic, a metrologist has to program it, set appropriate bases, construct supports, etc. As a result one method may be better than another. The simplest possibility then is to copy or implement the better one, if possible.

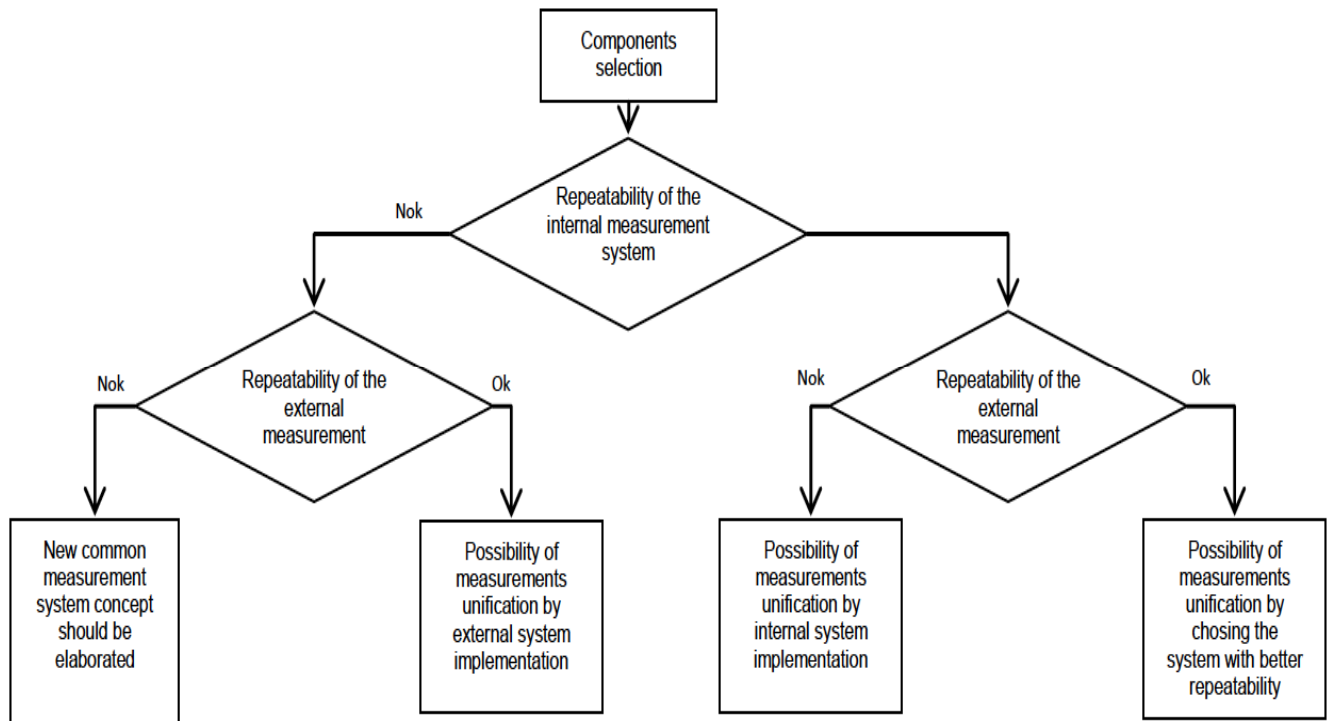


Figure 6: Possibilities of measurement systems unification.

How to choose? Similar approach to the one proposed for two systems comparison can be applied to characterize a system with better repeatability. After reiteration of each component measurement, size of the measurement error can be assessed with the usage of the regression method. Slightly different would be the interpretation of the term. The estimated value would mean the share of the repeatability error in the total observed variation from measurements.

It can happen that some existing constraints in equipment will make it impossible to be implemented in different environment (e.g. at supplier or customer). Then investments in device modifications can occur needed. Also, if none of the measuring method is repeatable enough, hard costs may become inevitable. On the other hand, it is possible to imagine many different methods enabling very good repeatability, but not reproducible when compared one to another. Although, such situation seems in the beginning to be comfortable, during trying to agree one common method, it appears difficult to finalize. Good management support in making decision is then welcomed.

Prior to assuring the relationship between systems, repeatability of both methods should be verified. Four milestones characterize the process of measurement system improvement:

- Choosing representative parts from the process
- Repeatability of both measurement systems
- Relationship between systems
- Improvements in measuring methods

In practice, the first step requires also deep understanding of the production process. In order to have good overview of the total variation it is necessary to draw a representative sample (sample variation should be close to the population). In order to assess if this requirement is fulfilled, historical production process should be analyzed (e.g. based on SPC measurements).

4. CONCLUDING REMARKS

In case of two measurement systems comparison, the regression method gives a straightforward answer for the question if the systems are comparable. Methods applied in the research are not really new, however their interpretation after some modifications become valuable especially for non-statisticians, who deal with similar problems.

The criterion of 10% and 30% for θ is proposed as an alternative for the approach from MSA norm. For the cases where two systems are needed to be

comparable, this method may be applied to verify results in very fast and easy way. Additionally, if some modification in measurement systems is implemented, then the analysis should be repeated. Naturally, the state of comparability between systems should be checked systematically (e.g. quarterly or yearly) to assure speaking the same language when using both measurement methods.

The thing needed to be underlined is that measurements cannot influence the measured feature or be destructive. Also components cannot change their characteristics in the period between measurements. The cost of the answer is half lower than using analysis of variance and more accurate. The reality shows that precise and automated measuring instruments like CMM has very good repeatability, however differences are still possible. Moreover, practice shows that the results of the analysis using regression model might be less influenced by the variance of parts in the sample.

Such approach may be applied also in comparison of different measuring instruments in order to answer if an expensive one is really necessary or the cheaper one can be also sufficient. When analyzing two systems where one belongs to supplier and another one is at customer, then such an analysis is essential to talk the same language when specifications of a product is touched.

REFERENCES

- [1] Taylor JR. Wstęp do analizy błędu pomiarowego. War saw 2012.
- [2] Chambers D, Wheeler D. Understanding Statistical Process Control. Kenoxville 2010.
- [3] Kończak G. Metody statystyczne w sterowaniu jakością produkcji. Katowice 2000.
- [4] Wheeler D. Zrozumieć zmienność. Klucz do zarządzania chaosem, Wrocław 2008.
- [5] Automotive Industry Action Group (AIAG). Measurement Systems Analysis Reference Manual. 4th edition. Chrysler Ford General Motors Supplier Quality Requirements Task Force 2010.
- [6] Dietrich E, Schulze A. Metody statystyczne w kwalifikacji środków pomiarowych maszyn i procesów produkcyjnych. Warsaw 2000.
- [7] Moatari AK. Design and Analysis of Gauge R&R Studies. Making Decisions Based on ANOVA Method. Engineering and Technology 2009.
- [8] Minitab, Inc. Minitab Statistical Software. Release 16 for Windows. State College. Pennsylvania 2009.
- [9] Altman DG, Bland JM. Measurement in Medicine: the Analysis of Method Comparison Studies. The Statistician 1983; 32: 307-317.
<http://dx.doi.org/10.2307/2987937>
- [10] Charles YT, Iglewicz B. Measurement Methods Comparison and Linear Statistical Relationship. Technometrics 1993; 41 (3): 192-201.
- [11] Carstensen B. Comparing and predicting between several methods of measurement. Biostatistics 2004; 5 (3): 399-413.
<http://dx.doi.org/10.1093/biostatistics/kxg043>
- [12] Linnet K. Evaluation of Regression Procedures for Methods Comparison Studies. Clin Chem 1993; 39 (3): 424-432.
- [13] Steiner SH, MacKay RJ, Ramberg JS. An Overview of the Shainin System for Quality Improvement. Quality Engineering 2007.
<http://dx.doi.org/10.1080/08982110701648125>

Received on 01-09-2014

Accepted on 21-10-2014

Published on 11-11-2014

<http://dx.doi.org/10.6000/1927-5129.2014.10.63>

© 2014 Kowal and Szymczak; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.