INFLUENCE OF THE INTERACTION BETWEEN PARTS AND APPRAISERS ON THE RESULTS OF REPEATABILITY AND REPRODUCIBILITY ANALYSIS

JIŘÍ PLURA, PAVEL KLAPUT

1   INTRODUCTION

The success of companies at the market has for several decades depended on the quality of the provided products and services. This quality cannot be achieved without functioning quality management system, whose main task is to plan, manage and continuously improve all the processes within the organization. All decision-making within the scope of this "trilogy of quality" should be done on the basis of collected data or facts. In the case of manufacturing processes, these facts represent the measured data of all the monitored quality parameters. An important condition for making the right decision is, in this case, a sufficient amount of quality data, provided thanks to a quality measurement system only. ISO/TS 16949 standard, which includes the requirements for quality management system in the automotive industry, says, in Clause 7.6.1, that statistical studies must be performed in order to analyse the variability of all types of measurement and test systems.

2   METHODOLOGY

In practice, these statistical studies are carried out according to several methods. One of them is the VDA 5 methodology, which originated in the German automotive industry. The basic principle of this methodology is the narrowing of tolerance of a given quality characteristic by the calculated measurement uncertainty. However, this methodology is used very rarely, and evaluation of measurement uncertainty is practically done only in testing and calibration laboratories. The most widely used methodology, utilized not only in the automotive industry, is the MSA methodology - Measurement System Analysis which was created by three American carmakers: Chrysler Group LLC, Ford Motor Company and Generals Motors Corporation. Its main principle is the evaluation of the most important statistical properties of the measurement systems, including stability, bias, linearity, repeatability and reproducibility. The most often performed of the studies is the study of combined repeatability and

reproducibility (GRR) of measurements. MSA Handbook (AIAG, 2010) describes three methods used for evaluating these studies. They are:

- Range method
- Average and range method
- Analysis of variance (ANOVA).

Using each of these methods brings certain advantages and disadvantages that can affect the quality and informative value of the results achieved. The range method, also referred to as "short method", is not normally used for verification of the measurement systems quality, but it serves for quick verification whether the percentage share of combined repeatability and reproducibility in total variation (% GRR) is satisfactory. Its major disadvantage is the fact that it does not allow independent evaluation of repeatability and reproducibility of the measurements, which is why this work will be focused only on the remaining two methods. The variability of results achieved using the average and range method is analysed in detail in previous work (Klaput & Plura, 2012). The following part of this work uses real and purposefully modified data in order to compare the GRR studies by means of the average and range method (A&R) and the ANOVA method. Based on the results of these studies, the conditions under which the results of the methods are going to be the same or, on the contrary, completely different are discussed.

2.1 Average and range method

The average and range method (A&R) is most commonly used for measurement system repeatability and reproducibility assessment in practice. The required data are obtained by repeated measurements of product samples realised by various appraisers. It uses a defined procedure, which includes both numeric and graphical evaluation of repeatability (EV) and reproducibility (AV). On the basis of their values, it is possible to calculate the combined repeatability and reproducibility (GRR) according to the relation (1).

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

The percentage share of GRR in the total variation and the number of distinct categories (*ndc*) are used as the criteria of the measurement system acceptability. They are calculated using relations (2) and (3).

$$\%GRR = \frac{GRR}{TV} . 100 \tag{2}$$

$$ndc = 1.41 . \frac{PV}{GRR} \tag{3}$$

where:

TV is total variation, $TV = \sqrt{(GRR)^2 + (PV^2)}$ \hfill (4)

PV – parts variation.

A measurement system is considered as fully acceptable in the cases, when %GRR value is lower than 10% and, at the same time, *ndc* value is at least 5.

2.2 ANOVA

The last, fourth edition of the MSA manual lays more and more stress on the evaluation of repeatability and reproducibility using the analysis of variance (ANOVA). As far as this method is concerned, you can divide the total variation into repeatability (EV), reproducibility (AV), parts variation (PV), and the interaction between appraisers and parts (INT). The GRR study using this method makes possible to obtain more information than in case of the average and range method, because it also provides information on how much of the total variation is caused by the interaction among the individual appraisers and parts. If this interaction is statistically significant, its value is presented separately, and combined repeatability and reproducibility is calculated as follows:

$$GRR = \sqrt{(EV)^2 + (AV)^2 + (INT)^2} \tag{5}$$

If the interaction is not statistically significant, it is assigned to the value of repeatability. That is how ANOVA method can detect more accurate estimates of the variances, provided that the measurement errors are normally distributed. This assumption can be assessed using appropriate graphical tools (Klaput & Plura, 2011). The disadvantage of this method is in more complicated calculations of the individual components of variability, and its application requires the use of a computer (Petrík & Palfy, 2011).

3   IMPACT OF CHANGES OF THE MEASURED VALUES ON THE RESULTS ACHIEVED BY VARIOUS METHODS

As already mentioned above, the results of GRR analysis obtained using both methods can be very different. This difference may be caused by the occurrence of a statistically significant interaction between the measured parts and appraisers (Osma, 2011; Kazerouni, 2009). In this article, we are going to focus on exploring the impact of outliers that simulates the effect of interaction between parts and appraisers. The outcomes of the analysis of repeatability and reproducibility obtained by the average and range method and ANOVA method were compared on real data of nuts height measurement, performed by three appraisers from Tab. 1 (Plura, 2001).

In order to analyse the partial results of these analyses as well, an application in MS Excel 2010 were prepared for both methods. The accuracy of the results was verified using Minitab 16 program. The obtained results are shown in the first line of Table 3. A comparison of the determined %EV, %AV, %GRR and *ndc* values clearly show minimum differences among the results of the applications of the individual methods. It is mainly related to the fact, that the variability caused by the interaction between the parts and appraisers was evaluated by ANOVA method as statistically insignificant (it is considered to be zero).

Table 1 Measured data of nuts height, mm (Plura, 2001).

Appraiser	Trials	Measured part									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	1	10.96	10.9	10.67	10.35	10.71	10.82	10.55	10.65	10.46	10.55
	2	10.97	10.89	10.68	10.4	10.68	10.81	10.56	10.63	10.47	10.57
B	1	10.99	10.89	10.68	10.42	10.73	10.8	10.54	10.64	10.45	10.57
	2	10.96	10.93	10.74	10.39	10.72	10.84	10.6	10.69	10.43	10.54
C	1	10.94	10.85	10.71	10.36	10.73	10.75	10.48	10.65	10.42	10.55
	2	10.91	10.84	10.64	10.33	10.65	10.76	10.49	10.66	10.45	10.53

The following solution stage deals with a simulation of the effect of increasing variability caused by the occurrence of outliers, which simulate the interactions between parts and appraisers, on the results obtained by both methods. That is why the measured values of one or two selected parts (both measurements for each part) were successively changed in case of appraiser A, while always maintaining the range of repeated measurements. The measured values of the selected parts were gradually increased or decreased by multiples of the standard deviation of repeatability, which was set to 0.025 mm. The changes of the measured values for all three cases are shown in Figure 1.

3.1 Simulation 1

In the first case, parts No. 3 and No. 8, which original measured values were close to the average value of all the measurements performed by the appraiser, were selected for the given changes. The measured values were increased in one part and decreased in the other one, so there was no change in the total average or the change of the range of averages of all the measurements of the individual parts. This setting of the performed changes ensured stability of the values of %EV, %AV, %GRR and *ndc*, evaluated by means of the average and range method (see Table 2).

Table 2 shows the summary results of the analysis of repeatability and reproducibility obtained using the average and range method and ANOVA method, depending on the number of standard deviations, by which the measured values of parts 3 and 8 were increased and decreased. Whereas in the case of the average and range method the results remain constant, when the ANOVA method is used, there are considerable changes related mainly to the occurrence of the part – appraiser interaction.
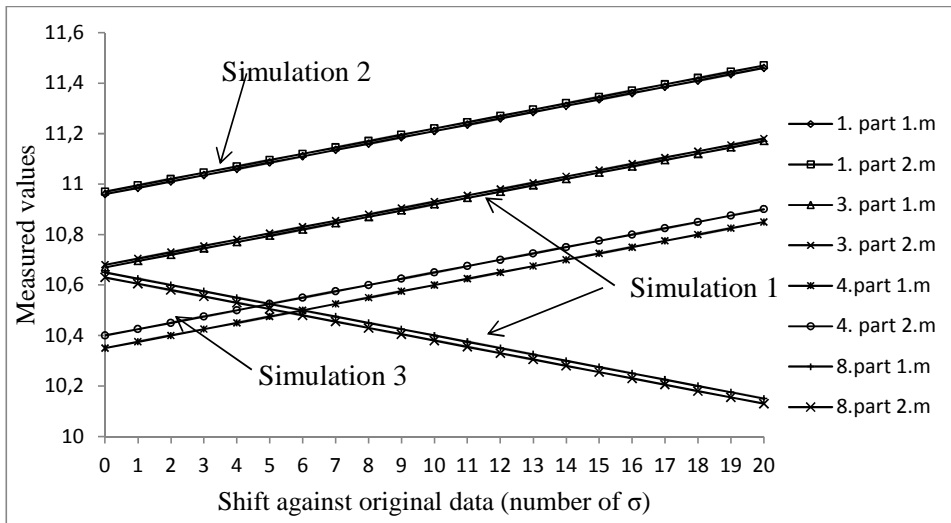
Figure 1 The changes of the measured values for all three simulations.

Table 2 Results of Simulation 1 for A&R and ANOVA methods.

Shift	A&R					ANOVA					
σ	%EV	%AV	%GRR	%PV	ndc	%EV	%AV	%INT	%GRR	%PV	ndc
0	13.55	11.62	17.85	98.39	7.77	11.14	12.93		17.09	98.53	8.13
1	13.55	11.62	17.85	98.39	7.77	11.33	12.91		17.43	98.47	7.96
2	13.55	11.62	17.85	98.39	7.77	10.85	12.88	7.41	18.39	98.29	7.53
3	13.55	11.62	17.85	98.39	7.77	10.50	12.83	10.94	19.86	98.01	6.96
4	13.55	11.62	17.85	98.39	7.77	10.01	12.76	14.44	21.72	97.61	6.34
5	13.55	11.62	17.85	98.39	7.77	9.36	12.69	17.89	23.85	97.11	5.74
6	13.55	11.62	17.85	98.39	7.77	8.52	12.60	21.28	26.16	96.52	5.20
7	13.55	11.62	17.85	98.39	7.77	7.45	12.50	24.59	28.58	95.83	4.73
8	13.55	11.62	17.85	98.39	7.77	6.04	12.39	27.83	31.05	95.06	4.32
9	13.55	11.62	17.85	98.39	7.77	3.97	12.27	30.97	33.55	94.20	3.96
10	13.55	11.62	17.85	98.39	7.77	0.00	12.13	34.02	36.12	93.25	3.64
11	13.55	11.62	17.85	98.39	7.77	0.00	11.98	36.93	38.82	92.16	3.35
12	13.55	11.62	17.85	98.39	7.77	0.00	11.82	39.73	41.45	91.00	3.10
13	13.55	11.62	17.85	98.39	7.77	0.00	11.65	42.43	44.00	89.80	2.88
14	13.55	11.62	17.85	98.39	7.77	0.00	11.48	45.01	46.45	88.56	2.69
15	13.55	11.62	17.85	98.39	7.77	0.00	11.31	47.47	48.80	87.28	2.52
16	13.55	11.62	17.85	98.39	7.77	0.00	11.13	49.83	51.06	85.98	2.37
17	13.55	11.62	17.85	98.39	7.77	0.00	10.95	52.08	53.22	84.66	2.24
18	13.55	11.62	17.85	98.39	7.77	0.00	10.77	54.22	55.28	83.33	2.13
19	13.55	11.62	17.85	98.39	7.77	0.00	10.59	56.26	57.24	81.99	2.02
20	13.55	11.62	17.85	98.39	7.77	0.00	10.40	58.19	59.12	80.66	1.92

When you change the values of the parts in question by only one standard deviation, the interaction is still statistically insignificant and the results remain

practically the same. When you make changes by two standard deviations and more, however, the value of interaction significantly increases. This also causes a significant increase of %GRR (see Figure 2). The behaviour of the values of repeatability (% EV) and reproducibility (% AV) is interesting. While the value of %AV decreases with increasing number of standard deviations and when the values are changed by ten standard deviations, it reaches zero, the value of %EV decreases only slightly. With the increasing shift of the measured values, there is also a slight decrease in %PV. Increase of %GRR and a slight decrease in %PV are reflected in a significant decline in the value of *ndc* (see Table 2). Even when you change the values by 7 standard deviations, the *ndc* value decreases below 5 and the measurement system would be classified as unacceptable.
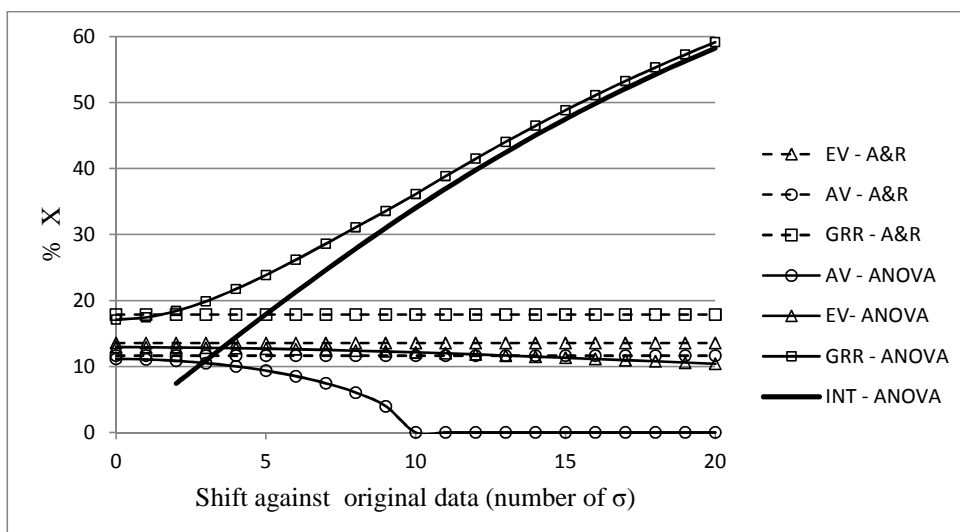

Figure 2 Changes in GRR study for Simulation 1.

The changes of the evaluated indicators are also connected with the change of the total variation (TV) to which the percentages of calculated indicators are related. It was calculated on the basis of the measured values, as the set of measured nuts represented the production range. While with using average and range method total variation did not change, the total variation calculated using the ANOVA method was increasing with the growing shift of values (see Figure 3), which somewhat mitigated the changes of the evaluated indicators.

3.2 Simulation 2

In the second case, the measured values of part No.1, which has the highest average value of all the measurements of all the measured parts, were increased. These changes therefore led to a change of the overall average of all the measurements of the given appraiser, but there was no change of the average range. The values of the final parameters for the second case are shown in Table 3 and illustrated in Figure 4.
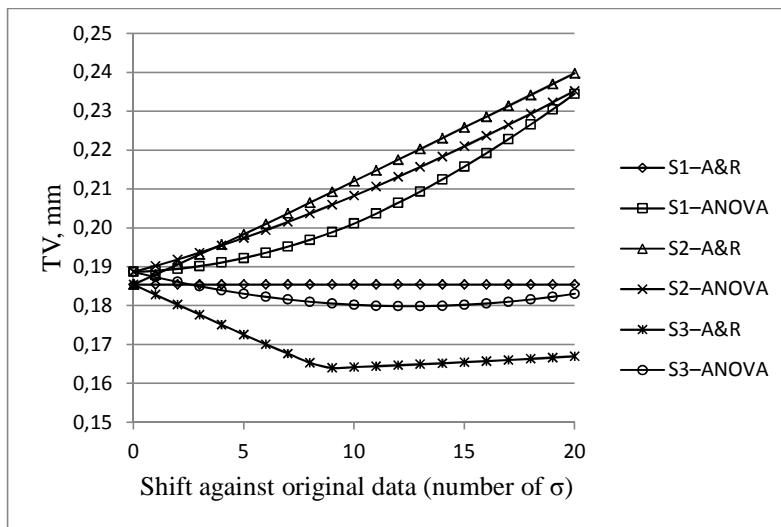
Figure 3 Changes of total variation in the individual simulations depending on the shift of measured data.

When applying average and range method, the growing shift of the measured values to higher values over the entire range was accompanied by a slight decrease in %EV, which is associated with an increase in the value of the total variation (see Figure 3), because the value of repeatability does not change, thanks to the constant value of the average range of the repeated measurements. In case of the percentage share of reproducibility (%AV), there is firstly a slight decline in the values and only with higher changes, there is an expected growth of values. The initial decreasing course is related both to increasing value of the total variation and also to the fact that with smaller changes of the values, the average of all the measurements of the given appraiser does not affect the value of the range of averages of all the measurements performed by the individual appraisers. The course %GRR practically copies the course of %AV.

When using the ANOVA method, the value of %GRR increases even with the smallest shift by one σ. In this case, however, the interaction part - appraiser itself is not statistically significant yet, and therefore the value of this interaction is included in the value of repeatability, which is reflected in a slightly higher value of %EV. If the shift of both measured values of the given appraiser is 2σ or higher, the interaction is already evaluated as statistically significant and hence its contribution is calculated independently. Increasing the size of interaction leads to a gradual reduction in the value of %AV and %EV, which is in line with the calculating relations of ANOVA method (Burdick, Borror & Montgomery, 2005).

Table 3 Results of Simulation 2 for A&R and ANOVA methods.

Shift	A&R					ANOVA					
σ	%EV	%AV	%GRR	%PV	ndc	%EV	%AV	%INT	%GRR	%PV	ndc
0	13.55	11.62	17.85	98.39	7.77	12.93	11.14		17.09	98.53	8.13
1	13.36	11.44	17.59	98.44	7.89	13.12	11.22		17.26	98.5	8.05
2	13.18	11.29	17.35	98.48	8.00	12.71	11.24	5.53	17.85	98.39	7.77
3	13	11.14	17.12	98.52	8.11	12.60	11.27	7.84	18.63	98.25	7.44
4	12.83	10.99	16.89	98.56	8.23	12.48	11.30	10.1	19.63	98.05	7.04
5	12.66	10.85	16.67	98.6	8.34	12.36	11.32	12.31	20.8	97.81	6.63
6	12.49	11.11	16.72	98.59	8.31	12.23	11.34	14.48	22.09	97.53	6.23
7	12.32	11.62	16.94	98.56	8.20	12.11	11.35	16.59	23.46	97.21	5.84
8	12.16	12.11	17.16	98.52	8.10	11.98	11.35	18.65	24.9	96.85	5.48
9	12	12.59	17.4	98.48	7.98	11.85	11.34	20.67	26.38	96.46	5.16
10	11.85	13.06	17.63	98.43	7.87	11.71	11.34	22.63	27.89	96.03	4.85
11	11.69	13.51	17.87	98.39	7.76	11.58	11.32	24.54	29.4	95.58	4.58
12	11.54	13.95	18.11	98.35	7.66	11.45	11.30	26.39	30.91	95.11	4.34
13	11.4	14.38	18.35	98.3	7.55	11.31	11.28	28.2	32.41	94.6	4.12
14	11.26	14.8	18.59	98.26	7.45	11.18	11.26	29.95	33.89	94.08	3.91
15	11.12	15.2	18.84	98.21	7.35	11.04	11.23	31.66	35.35	93.54	3.73
16	10.98	15.6	19.08	98.16	7.25	10.91	11.19	33.31	36.79	92.99	3.56
17	10.85	15.98	19.32	98.12	7.16	10.77	11.16	34.91	38.2	92.42	3.41
18	10.72	16.36	19.56	98.07	7.07	10.64	11.12	36.47	39.58	91.83	3.27
19	10.6	16.72	19.8	98.02	6.98	10.50	11.08	37.98	40.93	91.24	3.14
20	10.47	17.08	20.04	97.97	6.89	10.37	11.03	39.44	42.25	90.64	3.02

Table 3 and Figure 4 clearly show that the ANOVA method is much more sensitive in terms of the occurrence of interaction than the average and range method. Using the A&R method would, in this case, not change the evaluation of the acceptability of the measurement systems, not even for shift of 20σ. On the contrary, the evaluation of GRR using ANOVA method would, in terms of %GRR, rate the measurement system as unacceptable, even with a shift by 12σ.

With regards to the size of the achieved percentage interaction share (% INT) or %GRR, the values in question are lower in comparison to Simulation 1. This is caused by the fact that, in this case, there was a change of the two measured values, while in Simulation 1, four measured values were changed.

The differences in results obtained when using the individual methods also have an impact on the evaluation of the acceptability of the measurement system. When using the average and range method, the system of measurement over the entire range of simulated changes would remain acceptable, and when using the ANOVA method, it would become unacceptable with a shift of values as low as by 10σ, thanks to the low value of *ndc*.
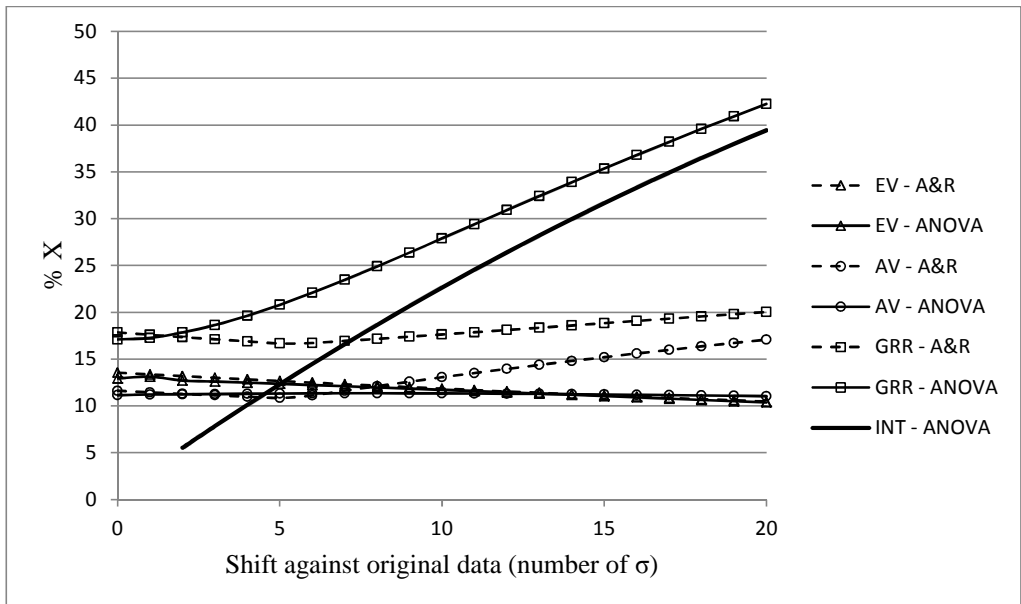
Figure 4 Changes in GRR study results for Simulation 2.

3.3 Simulation 3

In the third case, the measured values of part No. 4, which has the smallest average value of all the measurements of all the measured parts, from the same appraiser were gradually increased. As in the case of Simulation 2, there was a change of overall average of all the measurements of the given appraiser, but there was no change in the average range of the repeated measurements. The determined values of the final indicators for this case are shown in Table 4 and illustrated in Figure 5.

Because the simulation of occurrence of interaction in this case is very similar to the Simulation 2, one would expect that the final values or their changes depending on the size of the shift of two measured values of the given part will be similar, if not the same. However, when the average and range method was used, the first differences are apparent as early as in the course of dependence of %EV. While the percentage share of repeatability in Simulation 2 was decreasing over the entire range, in this case, small values of shift first lead to increase of this value, which remains constant when reaching the shift by approximately 9σ. The process is clearly related to the change of the total variation, whose dependence also has a shape of a broken curve (see Figure 3). The value of total variation TV decreases with small values of shift, because the variation range of averages of all the measurements of the individual parts used as the basis for calculation of variability between the measured parts (PV) declines. At the moment when the average of all the measurements of the given part reaches the level corresponding to the second smallest part, the range of the averages of the parts remains constant.

Table 4 Results of Simulation 3 for A&R and ANOVA methods.

Shift	A&R					ANOVA					
σ	%EV	%AV	%GRR	%PV	ndc	%EV	%AV	%INT	%GRR	%PV	ndc
0	13.55	11.62	17.85	98.39	7.78	12.93	11.14		17.09	98.53	8.13
1	13.73	11.77	18.08	98.35	7.67	13.10	11.41		17.37	98.48	7.99
2	13.93	11.93	18.34	98.3	7.56	13.57	11.70		17.92	98.38	7.74
3	14.13	12.11	18.61	98.25	7.44	13.19	11.90	6.56	18.93	98.19	7.31
4	14.34	12.29	18.88	98.2	7.33	13.26	12.14	9.07	20.14	97.95	6.86
5	14.55	12.47	19.16	98.15	7.22	13.33	12.38	11.61	21.58	97.64	6.38
6	14.76	13.13	19.76	98.03	7.00	13.38	12.60	14.16	23.21	97.27	5.91
7	14.97	14.12	20.58	97.86	6.70	13.43	12.82	16.73	25.00	96.83	5.46
8	15.19	15.13	21.44	97.68	6.42	13.48	13.03	19.31	26.91	96.31	5.05
9	15.31	16.07	22.2	97.51	6.19	13.51	13.23	21.88	28.92	95.73	4.67
10	15.29	16.86	22.76	97.38	6.03	13.54	13.42	24.46	31.01	95.07	4.32
11	15.27	17.65	23.34	97.24	5.87	13.55	13.61	27.03	33.16	94.34	4.01
12	15.25	18.43	23.92	97.1	5.72	13.56	13.78	29.58	35.34	93.55	3.73
13	15.22	19.21	24.51	96.95	5.58	13.56	13.94	32.12	37.55	92.68	3.48
14	15.2	19.98	25.11	96.8	5.44	13.55	14.09	34.64	39.77	91.75	3.25
15	15.18	20.57	25.71	96.64	5.30	13.54	14.22	37.12	42.00	90.75	3.05
16	15.15	21.51	26.31	96.48	5.17	13.51	14.35	39.58	44.22	89.69	2.86
17	15.12	22.27	26.92	96.31	5.04	13.48	14.47	42.00	46.42	88.57	2.69
18	15.1	23.03	27.54	96.13	4.92	13.43	14.57	44.38	48.61	87.39	2.53
19	15.07	23.78	28.15	95.96	4.81	13.38	14.66	46.72	50.76	86.16	2.39
20	15.04	24.52	28.77	95.77	4.69	13.33	14.75	49.01	52.89	84.87	2.26

A similar effect can be seen in the percentage share of reproducibility (%AV). The initial increase of this value (up to app. 6σ) is related, as in the case of %EV, to decreasing value of the total variation. The following, more significant increase is caused by the fact that the average of all the measurements of the given appraiser becomes the maximum value of the averages of all the measurements performed by the individual appraisers, which directly affects the range of averages.

When applying the ANOVA method, the values of %EV remain practically unchanged with increasing shift of the measured values, and %AV shows slow increase only. There is, however, a significant increase in the value of interaction between parts and appraisers, the course of which is then copied by the value of the percentage share of combined repeatability and reproducibility (%GRR). The analysis using ANOVA method leads to a change of the evaluation of acceptability of the measurement systems in this case as well. When the average and range method was used, the measurement system becomes unacceptable only when shifting the value by 18σ, which is caused by low value of *ndc*. The use of ANOVA method makes the system unacceptable for the same reason with the shift by 9σ.
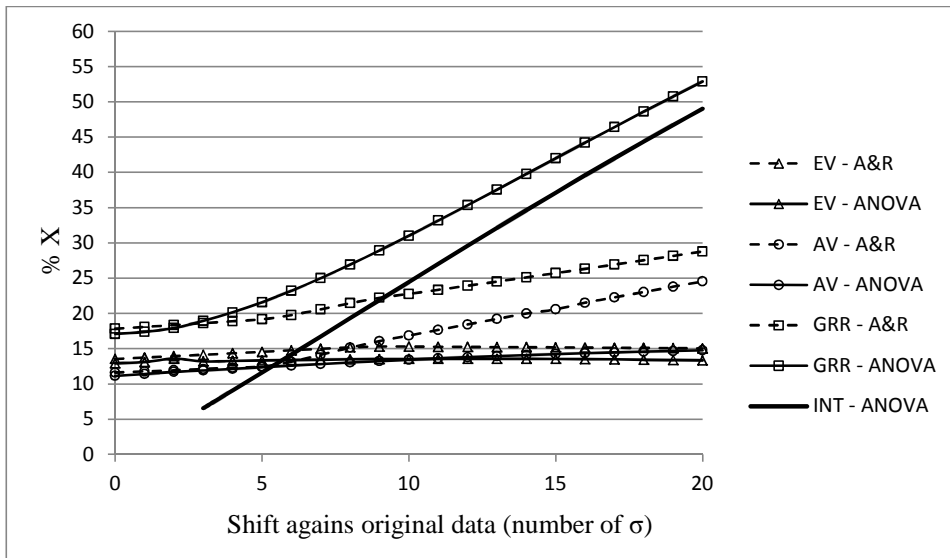
Figure 5 Changes in GRR study results for Simulation 3.

4   CONCLUSION

The results of the simulations show that the ANOVA method is more suitable for analysis of repeatability and reproducibility of the measurement system. Its main advantage is the ability to detect eventual interactions between parts and appraisers, which may significantly worsen the variability of the used measurement system. This makes the analyses using this method usually more sensitive to the occurrence of unusual situations, such as outliers.

In case of studies of repeatability and reproducibility of measurement systems, where these interactions do not occur, comparable results are achieved by means of the average and range method, whose undisputable advantage is the fact that the used procedure of evaluation is much more transparent and a series of partial results can be analysed as well. However, this method does not allow detecting variability caused by the interaction between parts and appraisers.

The analyses of the measurement systems based on the numerical evaluation must always be completed with appropriate graphic tools (Klaput & Plura, 2011). They will make possible to obtain a much more complex picture of the quality of the evaluated measurement system and to identify concrete causes of the measurement system properties.

ACKNOWLEDGEMENT

Engineering, VŠB-TU Ostrava with the support of Ministry of Education, Youth and Sports, Czech Republic.

REFERENCES

AIAG Reference manual, (2010), *Measurement Systems Analysis. MSA. 4th Edition.*, AIAG, Southfield.

Burdick, R. K, Borror, C. M, Montgomery, D. C. (2005), *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*, Cambridge University Press, Cambridge.

Kazerouni, A, M. (2009), "Design and Analysis of Gauge R&R Studies: Making Decisions Based on ANOVA Method", in *Proceedings of World Academy of Science, Engineering and Technology,* Vol. 40 No.4, pp. 31 – 35

Klaput, P., Plura, J. (2011), "The Importance of Graphical Tools for Measurement Systems Analysis.", in *Proceedings of QMOD Conference on Quality and Service Sciences 2011*, Servicios de Publicaciones Universidad de Navarra, Pamplona, Spain, 29-31 August, 17 p.

Klaput, P., Plura, J. (2012), "Influence of Used Method on the Results of Gage Repeatability and Reproducibility Study", in *Proceedings of Metal 2012 21st International conference on Metallurgy and Materials*, Tanger, Brno, Czech Republic, 23 – 25 May, 6 p.

Osma, A. (2011), "An assessment of the robustness of gauge repeatability and reproducibility analysis in automotive components", *Journal of Automotive Engineering*, Vol. 225, Issue 7, pp. 895 – 912

Petrík, P., Palfy, P. (2011), "The Influence of the Load on the Hardness", *Metrology and Measurement Systems*, Vol. 18, No. 2, pp. 223-234.

Plura, J. (2001), *Quality Planning and Continuous Quality Improvement. 1st edition* (in Czech). Computer Press, Prague.

ABOUT THE AUTHORS

Prof. Ing. **Jiří Plura**, Ph.D., VŠB-Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Department of Quality Management, 17.listopadu, Ostrava-Poruba 708 33, Czech Republic,
e-mail: jiri.plura@vsb.cz.

Ing. **Pavel Klaput**, VŠB-Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Department of Quality Management, 17.listopadu, Ostrava-Poruba 708 33, Czech Republic, e-mail: pavel.klaput@vsb.cz.