

Pavĺina
Mikulov¹
Andrs C. Garca

APPROACHES TO REPEATABILITY AND REPRODUCIBILITY ANALYSIS OF NON- REPLICABLE MEASUREMENT SYSTEMS

Abstract: *Measurement systems have to deal with processes in which, by the nature of the measured object or by the type of measurement itself, measurements are not repeatable in the sense that each sample unit can be measured only once. Such situations are usually referred as non-replicable measurement systems. The paper presents the approaches to GRR analysis in these cases. On the base of current development in this branch, the special attention is focused on the appropriate setup of GRR analysis and the choice which type of GRR design to use.*

The aim of the paper is to map out the current approaches to GRR analysis of non-replicable measurement systems with all the assumptions as for consistency, bias and temporal stability of objects. Subsequently, it is planned to verify these approaches in practice what is the opportunity to suggest the improving actions to obtain the results of high quality.

Keywords: *Measurement system, Repeatability, Reproducibility, Non-replicable measurements, GRR*

1. Introduction

The quality of measurement system is influenced both by the measuring instrument and the measurement system conditions, and the factors including the appraisers who carry out the measurements. The measurement system requires attention and monitoring not to provide distorted information which can lead to wrong decisions. The suitability of a measurement system is an important part of quality planning and quality improvement (Plura, J., 2001), and it is even strictly required in automotive industry (AIAG, 2006). A set of methods known as MSA (Measurement System Analysis) is used to evaluate a measurement system capability to provide the data of the highest quality. It can prevent the problem of bad parts to be accepted or good parts to be rejected (in the case of attribute measurement) or the lack of accuracy and precision (in the case of variable measurement). There are various factors that affect the measurement systems and influence the measured values related to the measurement system

properties. Therefore, analyzing these properties leads to the detection of the causes of low-quality data. The most used analysis of measurement system, which makes possible to evaluate repeatability and reproducibility of measurement system is GRR analysis.

In most cases, the readings of measurement systems can be replicated for each part. This means that parts or samples are not altered by the measurement process, so measurements can be repeated providing comparable data. However, not all measurement systems have this characteristic. This is the case of non-replicable measurement systems where the part is altered or even destroyed during the measurement. In statistics, a replicable measurement system is one where any given part may be measured multiple times by the same or different appraisers and the result obtained falls within a predictable range of values (Miner, G., 2016).

One of the essential properties of measurement system is its consistency. A consistent measurement process is in statistical control with respect to variation. Moreover, bias and stability

¹ Corresponding author: Pavĺina Mikulov
Email: pavlina.mikulova.st@vsb.cz

of the measurement system is necessary to meet. The bias refers to the location of the data relative to a reference value; presents the difference between the observed average of measurements and the reference value. The stability is the change in bias over time. A stable measurement process is in statistical control with respect to location.

2. Non-replicable measurement systems

As mentioned, normally in measurement systems, the readings can be replicated for each part. Though, this is not the case of non-replicable measurement systems where the part is destroyed or its characteristic is changed during measurement. Following are the general categories of non-replicable measurement systems (MSA Work Group, 2010):

- Destructive measurement systems,
- Systems where the part changes during measurement process.

In this case, different appraisers cannot measure the same part nor each appraiser obtains successive measurements of this part. This restriction impacts the experimental design to be used as well as the GRR analysis approach (Miner, G., 2016). Two approaches are usually considered:

- Crossed GRR analysis,
- Nested GRR analysis.

Software tools, as Minitab, offer these options to practitioners. Section 3.1 deals with this question in more detail.

2.1. Destructive measurement systems

In terms of destructive measurement, the part is changed or destroyed during the measurement process. Thus, different appraisers cannot measure the same parts neither repeated readings cannot be taken on any single part, e.g. destructive weld testing, destructive plating tests, salt spray or humidity booth testing, impact testing (gravimeter) or mass spectroscopy and other material characteristic testing processes (Donald, S. E., 2006).

Following are the advantages of destructive measurements (LTI Group, 2019):

- ✓ Allow identifying the mechanical properties of material (fracture strength, elongation, modulus of elasticity etc.),
- ✓ Predict the approximate nature of failure or breakdown that may occur during the lifetime of the product use,
- ✓ Verify e.g. the surface preparation, curing conditions, working conditions and system products.

Furthermore, the destructive measurements have some disadvantages must be mentioned:

- × Costs and time-consuming method,
- × Harder to ensure the consistent measurement process (a sample cannot be re-measured once have been tested.)
- × Provide limited options of analyzing the obtained data (see section 3.1.).

There are some cases where the part itself might be destroyed, but the measurement is still replicable. This may occur if the part must be damaged to gain access to the characteristic to be measured. Once access is gained the characteristic may be measured repeatedly.

2.2. Systems where the part changes during measurement

There are other non-replicable measurement systems where the part itself is not destroyed by the measurement process but the characteristic being measured changes, e.g. Leak tests with qualitative data, testing using engine test stands, transmission test stands, vehicle dynamometers, torque measurement etc. (MSA Work Group, 2010).

According to (Miner, G., 2016) these systems are divided to the follows, when:

- Part changes during measurement process (rubber parts soften when flexed or harden when left alone; uncured rubber cures when tested in a rheometer).
- Measured characteristic changes over time (dimensions of freshly molded plastic parts, viscosity of materials with shelf lives)
- Part cannot be physically reintroduced to the measurements (in-line measurement devices)
- The part cannot be re-measured in the same location (e.g. hardness test).

Analysis of these systems depends on whether (MSA Work Group, 2010):

- a homogeneous set of parts (small between part variation) can be found to represent a single part,
- the shelf life of the characteristic is known and extends beyond the expected duration of the study (i.e. the measured characteristic does not change over the expected period of use),
- the dynamic (changing) properties can be stabilized.

On the base of current state in non-replicable measurement systems, it is necessary to focus on the possible analysis of the data obtained from this kind of measurements. Then, it is required to realize which is the appropriate way to stabilize the measurement process and make it consistent. The attention is focused on appropriate setup of GRR analysis considering the number of samples, appraisers and measurements, so that it is the compromise between the requirement for obtaining the highest quality results and the requirement for minimum financial and time-consuming performance of this analysis. The further important decision is which type of GRR designs to use from both described in section 3.1.

2.3. Types of sample preparation for non-replicable measurements

Now, having a non-replicable measurement system, the following approaches are recommended to use in order to perform optimal measurement system analysis and gain knowledge with similar results as the results obtained from such system with replicable measurements (Miner, G., 2016):

- ✓ Split samples – the parts or material collected to represent one part are split or subdivided into smaller units. The smaller units are used for the repeat trials and between appraiser trials (e.g. a steel bar is cut out for smaller parts). According to MSA reference manual fourth ed. (MSA Work Group, 2010), V3 and V4 study can be used based on this type of measurements.
- ✓ Consecutive samples – consecutive parts are used to represent one part for the repeat trials and between appraiser trials. This way is used when the parts cannot be subdivided and consecutive parts can reasonable be expected as homogeneous as if they occurred

in an auto-correlated process. Using consecutive parts, the V3a and V4a study are used following the MSA reference manual (MSA Work Group, 2010).

- ✓ Stabilized parts – parts or systems are stabilized before measurement process. This stabilization depends on the product and characteristic. Some characteristics may stabilize when pre-measured a number of times.
- ✓ Regression approach – when the change in the characteristic over time or activity is known and has a defined relationship [$Y=f(x)$] (e.g. shrinkage of plastic parts), the subsequent measurements are adjusted using this relationship and then analyzed.

3. Gage Repeatability and Reproducibility analysis

Since measurement systems are to be used in making decisions about products, processes, or services, an analytic conclusion about the measurement system is necessary. The transition from enumerative to analytic results requires subject matter knowledge and expertise to assure that all expected measurement sources of variation are considered in the design (MSA Work Group, 2010).

The realization of the measurement under the same conditions is impossible. The conditions are changeable what is mostly caused by the change of operator performing the measurement. In general cases, two major sources of measurement system variability are repeatability and reproducibility, see Figure 1 (Plura, J. 2012). Various approaches can be used to evaluate repeatability and reproducibility, the most used are e.g. Range method, Average and Range (A&R) method or ANOVA (MSA Work Group, 2010).

To approximate the measurement system variation, the method based on range can be used. The main advantage of Range method is the quick carrying out. However, the fact that it does not allow the individual components of variation to be distinguished – measurement system repeatability and reproducibility – is considered to be a disadvantage. A&R method uses subgroups of data to determine the variance due to the various sources of variation. Applying this method the variation induced by repeatability or

reproducibility can be distinguished. It does not take into account the operator-part interaction. ANOVA is a technique that examines what sources of variation have a significant impact on

the results. This approach actually adds another source of variation to the mix and it is able to identify the operator-part interactions (Plura, J., 2012).

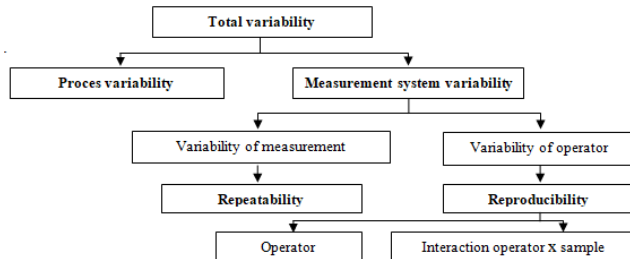


Figure 1. Total variability layout of measured data

Thus, GRR study is a designed experiment to study the variation in measurement results. The experiment is design to determine how much variation is due to the measurement method and how much is due to the appraisers. This is performed by measurements of parts from the process and analyzed by methods of GRR study.

3.1. Designs to assess the non-replicable measurement system

If the part is altered or destroyed during the

measurement process, you cannot just select parts from the process and have each appraiser measure each part. Hence at this point, working with the batches is necessary. Batch is considered to be one of the kinds described in section 2.3.

For assessing the non-replicable measurement systems two types of GRR design are known – crossed and nested GRR analysis. Which one of these to use is shown in Figure 2.

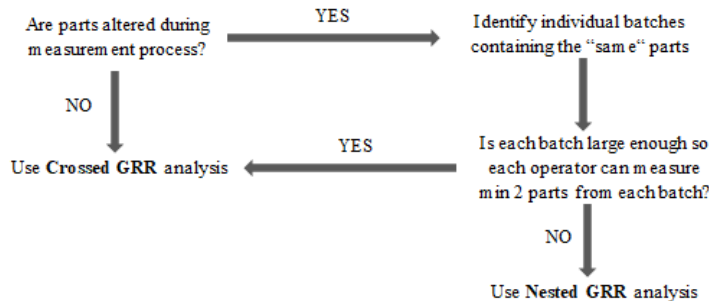


Figure 2. Selection of GRR design

As for crossed GRR design, conducting GRR analysis of non-replicable measurement system requires making a critical assumption: to identify a batch of parts where the parts are so close to being alike that these in the batch are the “same“. Thus, the assumption is that the batch is homogeneous. This means measuring any part of the batch for the same characteristic, the results should be similar with a small sign of variation due to repeatability same as it occurs in the

replicable measurement system. But unlike the crossed design of replicable measurement systems, the same part is not really being measured in spite of the parts are the “same“ (come from a homogeneous batch). In reality, the variation is due to the measurement system repeatability. If a large amount of variation occurs, then the question is whether the batch is really homogeneous (McNeese, B., 2016).

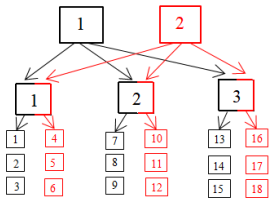


Figure 3. Crossed GRR design (11)

In nested GRR design, in comparison with the crossed one, the difference occurs at the stage when determining if there are sufficient parts from each batch for each operator to measure. If there are, as described above, the crossed GRR design is used. However, if there are not sufficient parts in each batch for each operator to measure multiple times, then it is necessary to use a nested GRR analysis.

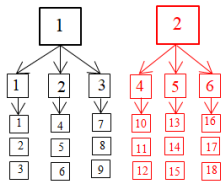


Figure 4. Nested GRR design (11)

De Mast and Trip (De Mast et al., 2005) claims that the option of the nested design may result in misunderstanding. Under the assumption of objects homogeneity the crossed design could be used, but with repetitive measurements on single objects replaced with measurements on different objects. The appraiser selects I parts of JK objects each, which are assumed to be homogeneous. Each of J operators measures K of these objects once. The effect of the JK objects of a sample is nested within the operators factor, which might be the reason of that so many courses refer to this design as being nested. Although, the samples of the batch which are measured by the operator are nested within their combination, in fact, the batches are crossed with operators. The name of Nested is somehow deceptive since the samples factor is still crossed with the operators factor, whereas the objects factor – though it is indeed nested – does not show up in the process. According to (McNeese, B., 2016), ANOVA allows to analyze the results of a nested GRR design provided by software Minitab or Statgraphics and it really does, but, it is claimed A&R method cannot be used. Let's see the

assumptions of non-replicable measurements and, subsequently, we will try to verify this.

4. Assumptions for the proper GRR analysis application

One of the important questions is how to monitor the consistency of a non-replicable measurement system over time. In other words, how to ensure that the repeatability of the non-replicable measurement is the “same” over time. This is simple to do for replicable measurement systems. Using a standard or a reference part, we can use a \bar{X} and R control chart (if the standard is measured $K \geq 2$ times in each cycle) or Individual \bar{X} and moving R chart (if the standard is measured once in each cycle). From the \bar{X} or individual \bar{X} chart the measurement process stability (bias constancy over time) can be controlled. From R or moving R chart the measurement process consistency (repeatability constancy over time) can be controlled. The way how to analyze consistency of a non-replicable measurement is to use the approach of samples multiplication. These are simply multiplied samples in lots of K which are the “same”. They are not identical, but they are as close as possible to being alike. The range represents the variation in the method of non-replicable measurement. The variation of non-replicable measurement consistency is the variation in the results between K samples. Generally, the consistency is considered to be the degree of change of repeatability over time (McNeese, B., 2015).

Reference Manual MSA fourth edition presents S4 and S4a study used in this case, where S4 represents the study of Split samples (General) or Single sample per cycle, and S4a study is the same as S4 with homogenous parts from different lots, it is an upper bound study. Further, the assumptions for temporal stability of objects should be met that means it does not matter at which time the samples are measured. If this is not met, there are some alternatives which can be used, see (De Mast, J., Trip, A., 2005).

The nested GRR design provided by software tool Minitab is not suitable for GRR analysis due to the deceptive effects of appraisers factor and differences between batch groups assigned to them. The obstacle occurs when comparing the differences between appraisers. The crossed GRR design is preferred, though, it signifies to ensure

the sufficient amount of samples to being measured. Thus, the planned experiment presented in chapter 6 will deal with crossed GRR design.

5. Determination of variance components by GRR analysis using crossed design

Meeting the assumptions is the first essential step to ensure a well-working measurement system. Further important point is to be aware of the factors being investigated – part and operator. In general cases, a complete factorial design with replications includes I parts and J operators, each operator measures each of I parts r -times. Considering the recommended procedure for operators to proceed to the next measurement of the same part after all I parts have been measured, it is appropriate to consider the layout of blocks made up of individual replicas (Jarošová, E., 2018). The model has following form:

$$y_{ijk} = \mu + p_i + o_j + (po)_{ij} + b_k + e_{ijk}$$

$$i=1,2,\dots,I; j=1,2,\dots,J; k=1,2,\dots,r$$

where μ is the overall mean, p_i is the effect of part i , o_j is the effect of operator j , $(po)_{ij}$ is the effect of interaction between part i and operator j , b_k mean random effect of block k and e_{ijk} represents the random element. Normally, it is assumed the results are not influenced by the block (measuring time), therefore the model without the blocking factor is used:

$$y_{ijk} = \mu + p_i + o_j + (po)_{ij} + e_{ijk}$$

The task is not to compare the particular parts or operators participating in the experiment, but to measure the variability between different parts and operators in general by assuming that p_i , o_j , $(po)_{ij}$, e_{ijk} are mutually independent and normally distributed random variables with zero means and variances σ_p^2 , σ_o^2 , σ_{po}^2 and σ^2 . Then, the variance of response Y can be demonstrated:

$$\sigma_t^2 = \sigma_p^2 + \sigma_o^2 + \sigma_{po}^2 + \sigma^2$$

The aim of GRR analysis is to determine the components of variance due to the measurement system σ_{GRR}^2 and compare it with total variance (σ_t^2):

$$\sigma_{GRR}^2 = \sigma_o^2 + \sigma_{po}^2 + \sigma^2$$

where σ^2 represents repeatability (EV)² and $\sigma_o^2 + \sigma_{po}^2$ reproducibility (AV)².

Finally, considering GRR analysis for crossed design use, where parts with repeated measurements are replaced by batches of similar samples, slight changes will be made in the equations. The change will be in replacing “ p ” by “ b ”, ie. batches instead of parts:

$$y_{ijk} = \mu + b_i + o_j + (bo)_{ij} + e_{ijk}$$

Thus, the study consists of J operators and I batches of homogenous parts (each batch formed by $J \times r$ parts). The r blocks mean each operator measures once I part from each batch and each block.

Several methods for estimate the variance components exist. Best-known methods are ANOVA, ML (Maximum Likelihood) or REML (Restricted Maximum Likelihood). ANOVA estimators of all variance components are unbiased and have the smallest variance of all estimators which are both quadratic functions of the observations and unbiased. This is the case of balanced data (Searle et al., 1992). Assuming normal distribution, the estimators are minimum variance and unbiased. If the distribution of estimated variance components with exception of the estimate σ^2 cannot be described by any theoretical model, the exact confidence limits for σ_{GRR}^2 cannot be found. There are three methods presented in (Jarošová, E., 2018) for constructing approximate confidence limits which can be used. However, the disadvantage of ANOVA is that it can give negative estimates of variance components. From the statistical point of view, the way of estimate is not definite if some of the estimates are negative.

6. Discussion the planned measurements

The measurements to be performed for verifying the capability of non-replicable measurement system relate the process of measuring the oxygen volume in concrete blocks using two methods – Voltammetry (Van Brussel et al., 2003) and Impedance spectroscopy (Correia et al., 2006). In voltammetric methods, the range of potentials are scanned where the generated current is directly proportional to the concentration of electroactive species present in the sample (Mistry et al., 2014). Impedance spectroscopy measures the resistance and capacitance properties of a material. An impedance spectrum is obtained by varying frequency over a defined range. The capacitance

and resistance of the system can be then calculated by measurement of the in-phase and out-of-phase current responses (Correia et al., 2006).

For the experiment the four types of concrete is going to be used and measured by two operators. These types represent the batches. From each concrete type, the six homogenous samples are being prepared – three samples for each operator. Thus, we have 2 operators, 4 batches consist of 6 samples, ie. 24 samples in total, what allow each operator to measure 3 times (see Figure 5) and preserve the repeatability and reproducibility same as in the case of replicable measurement system with all the assumptions (meeting consistency, bias and temporal stability of homogenous samples).

About the types of concrete, four types will be measured as mentioned. These types (batches) differ in the ratio of water to cement (w/c), following are the ratios used for preparation of samples to be measured:

1. Batch consists of the samples with w/c ratio equal to 0.6.
2. Batch consists of the samples with w/c ratio equal to 0.5.
3. Batch consists of the samples with w/c ratio equal to 0.4.
4. Batch consists of the samples with w/c ratio equal to 0.5sf. A part of the cement is replaced by sf, ie. silica fume.

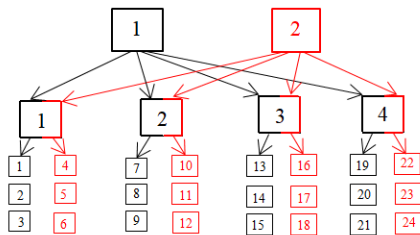


Figure 5. Crossed GRR design to be used for the system of measuring the oxygen volume

In terms of the measurement conditions, the operational pressure will be adjusted on 0 mbar and 1000 mbar (1 atm). In vacuum conditions the concentration of oxygen is equal to 0. In 1 atm conditions there is an oxygen mole fraction of 21% (atmospheric condition). This oxygen mole fraction will be changeable within the

measurement process by changing the partial pressure of oxygen. The electrodes of inox wires with the same working surface, the diameter of 0.08mm and length of 12cm will be used. Next, the temperature is considered to be equal to the lab temperature, ca. 20°C and the relative humidity to be 100% (Brownson et al., 2014).

For the evaluation of planned measurement system, the crossed GRR design will be used and compared with the most used methods – ANOVA and A&R method.

7. Conclusion

On the base of current development in the designs of GRR study for non-replicable measurement systems, the conclusion is that the nested GRR design provided by Minitab is not suitable for GRR analysis due to the delusive effects of operator’s factor and differences between batch groups assigned to them. The crossed GRR design is admittedly preferred, nevertheless, it is necessary to ensure the sufficient amount of measured samples. This will be verified in future work discussed in chapter 6 by using the described methods, ANOVA and A&R method, though, some of the authors stated in the paper do not recommend to use that. But, the information found seem to be contradictory and needed to be verified since there is still lack of evidence.

The GRR analysis for non-replicable measurement systems can provide an evidence of satisfactory measurement system if the total gage variance σ_{GRR}^2 is low. On the contrary, if the gage variance is high (the components are influenced by sample-to-sample variation), nothing can be concluded. In spite of the fact this situation of high total gage variance may occur, according to meeting the mentioned assumption, it is possible to ensure the consistent and stable measurement process as well as in the case of replicable measurement system. This is the sign to reach just the low total gage variance.

Acknowledgment: The work has been supported by the specific university research of the Ministry of Education, Youth and Sports of the Czech Republic No. SP2019/62 and elaborated in the framework of the grant programme „Support for Science and Research in the Moravia-Silesia Region 2018” (RRC/10/2018), financed from the budget of the Moravian-Silesian Region.

References:

- AIAG. (2006). *Production Part Approval Process (PPAP)*, Chrysler Gp. LLC, Ford Motor Company, General Motors Corp.
- Brownson, D.A.C., Banks, C.E. (2014). Interpreting Electrochemistry. *The Handbook of Graphene Electrochemistr*, 23-77. doi: 0.1007/978-1-4471-6428-9-2
- Brussel, V.M, Kokkinidis, G., Hubin, A., Buess-Herman, C. (2003). Oxygen reduction at platinum modified gold electrodes. *Electrochimica Acta*, 48, 3909 – 3919. doi: 10.1016/S0013-4686(03)00529-2
- Correia, M.J., Pereira, E.V., Salta, M.M., Fonseca, I.T.E. (2006). Sensor for oxygen evaluation in concrete. *Cement & Concrete Composites*, 6, 226-232. doi:10.1016/j.cemconcomp.2006.01.006
- De Mast, J., Trip, A. (2005). Gauge R&R Studies for Destructive Measurement. *Journal of Quality Technology: A Quarterly Journal of Methods, Applications and Related Topics*, 37:1, 40-49.
- Donald, S. E. (2006). In Quality Progress official publication of ASQ: *Appraiser Variation in Gage R&R Measurement*. Retrieved from: <http://asq.org/quality-progress/2006/05/measure-for-measure/appraiser-variation-in-gage-rr-measurement.html>
- Donald, S. E. (2006). In Quality Progress official publication of ASQ: *Improved Gage R&R Measurement Studies part one*, Retrieved from: <http://asq.org/quality-progress/2006/03/measure-for-measure/improved-gage-rr-measurement-studies.html>
- Jarošová, E. (2018). *Destructive R&R study – evaluation problems*. Paper presented at the Proceedings of the 17th Conference on Applied Mathematics APLIMAT 2018, Bratislava, Slovakia.
- LTI Group. (2019). In Laboratory Testing Inc. Materials testing, nondestructive testing & calibration services: *Destructive Testing: Providing Samples for Analytical Testing*. Retrieved from The Labtesting website: https://www.labtesting.com/wp-content/uploads/2012/12/company_testing_brochure.pdf
- McNeese, B. (2015). In BPI Consulting LLC: *Monitoring Destructive Test Methods*. Retrieved from The SPC for Excel website: <https://www.spcforexcel.com/knowledge/measurement-systems-analysis/monitoring-destructive-test-methods>
- McNeese, B. (2016). In BPI Consulting: *Gage R&R for Non-Destructive and Destructive Test Methods*. Retrieved from The SPC for Excel website: <https://www.spcforexcel.com/knowledge/measurement-systems-analysis/gage-rr-for-non-destructive-and-destructive-test-methods>
- Miner, G. (2016). In Miner's blog, *Intro to MSA of Continuous Data – Part 6: R&R for Non-Replicable Measurements*. Retrieved from: <http://www.qualityforumonline.com/forum/index.php?resources/intro-to-msa-of-continuous-data-%E2%80%93-part-6-r-r-for-non-replicable-measurements.40/>
- Mistry, K.K., Layek, K., Mahapatra, A., RoyChaudhuri, C., Saha., H. (2014). *A review on amperometric-type immunosensors based on screen-printed electrodes Analyst*, 139 (10), 2289-2311.
- MSA Work Group. (2010). *Measurement System Analysis, Reference Manual, fourth ed.* Chrysler Group LLC, Ford Motors Corporation. ISBN 978-1-60-534211-5
- Plura, J. (2001). *Quality Planning and Continuous Quality Improvement*, Praha, Czech republic, Computer Press, 244. ISBN 80-7226-543-1
- Plura, J. (2012). *Quality Planning II*, first ed. VSB-Technical university of Ostrava, Ostrava, Czech republic, 172. ISBN 978-80-248-2588-5
- Searle, S.R., Casella, G., Mc Gulloch, C.E. (1992). *Variance component*. John Wiley & Sons, New York, United States, 501. ISBN-I 3 978-0-470-00959-8

Pavlna Mikulová

VSB – Technical University
of Ostrava,
Ostrava,
Czech republic
pavlina.mikulova.st@vsb.cz

Andrés Carrión García

Universitat Politècnica de
València,
Valencia,
Spain
acarrión@eio.upv.es
