IMEKO TC11 & TC24 Joint Hybrid Conference
October 17-19, 2022, Dubrovnik, Croatia

DATA SCIENCE FOR TESTING, INSPECTION AND CERTIFICATION

M. Cundeva-Blajer

Ss. Cyril and Methodius University in Skopje, Faculty of Electrical Engineering and Information Technologies-Skopje, Republic of North Macedonia, mcundeva@feit.ukim.edu.mk

Abstract:
Decisions in conformity assessment, especially in the testing, inspection and certification (TIC) sector are predominantly based on the measurement data. The digital transformation has started to impact the TIC sector. The quality of the data in big data analytics, is not always sufficiently addressed, especially in sectors with traditionally empirical approaches, as TIC. This contribution conducts a survey of possibilities for application of data science in the TIC decision making processes, based on conclusions with complementary usage of experimental "measurements" and the "data science", with a case study in estimation of the instruments re-calibration interval with data fusion approach.

Keywords: data science, testing-inspection-certification, re-calibration interval, data fusion

1. INTRODUCTION INTO CLASSES OF DATA SCIENCE PROBLEMS IN TIC

Many decisions in conformity assessment, especially in testing, inspection and certification (TIC) are based on measurement data. Measurements are vital in environmental protection, healthcare, industry, trade, energy sector etc. The digital transformation has started to create significant impact on the TIC sector. Data is becoming more important in all sectors, as huge amount of data is produced, stored, processed, and utilized for increasing number of applications. Big data analytics, artificial intelligence, machine learning, internet of things etc. are based on data, but the quality of the data is an issue not always properly addressed, especially in sectors with traditions based on empirical approaches, like the TIC. Poor data will produce poor models, incorrect results and finally wrong decisions. Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data, and apply knowledge from data across a broad range of applications [1, 2]. The revival of measurement and data science is caused by the revolution of sensory devices and the emerging data transmission, storage and processing capabilities available and variously deployed. Due to the huge amount of recorded information and the theoretical results of measurement and data science, numerous newly developed products are invented, and smart services and support are enabled. This contribution conducts a survey of the possibilities for application of the latest data science achievements in the TIC decision making processes, based on conclusions with complementary usage of "measurements" as completely experimental and the "data science" as methodology oriented towards modelling and simulation, combined with high complementarity and synergy, i.e. the data fusion approach. The contemporary scientific methodologies require experimental verification, whenever possible, for sustaining the theory validity. Experimental proof comprises a measure either quantitative or non-quantitative, often called 'qualitative', of the observable quantities derived by means of measurement. The degree of consistency of different measurement results, gained by various independent experimenters or by the same experimenter at different times, is a measure of results reliability representing the observed quantity, considering that experimental knowledge is generally imperfect to some degree and the combination of observations are standard and essential practices, [3].

Several classes of data science problems for which techniques might be developed and evaluated across different domains in the TIC sector, are [1]:

- Detection -finding data of interest in given dataset.
- Anomaly detection - identification of system states that force additional pattern classes in a model. Outlier detection is associated with identifying potentially erroneous data items forcing changes in prediction models "influential observations".
- Cleaning - elimination of errors, omissions, and inconsistencies in data or across datasets.
- Alignment - relating different instances of the same object [4], e.g., a word with the corresponding visual object, or time stamps associated with two different time series. Data alignment is frequently used for entity resolution, identifying common entities among different data sources.
- Data fusion - different representations integration of the same real-world object, encoded in a well-defined knowledge base of entity types [5].

- Identification and classification - attempt to determine, for each item of interest, the type or class to which the item belongs [6].
- Regression - finding functional relationships between variables.
- Prediction - estimation of a variable or multiple variables of interest at future times.
- Structured prediction - tasks where the outputs are structured objects, rather than numeric values. A desirable technique to classify a variable in terms of a more complicated structure than producing discrete or real-number values.
- Knowledge base construction - construction of a database with a predefined schema, based on any number of diverse inputs.
- Density estimation - production of a probability density (distribution function), beside a label/value.
- Joint inference - joint optimization of predictors for different sub-problems using constraints that enforce global consistency used for detection and cleaning for more accurate results.

Data science involves ranking, clustering, and transcription ("structured prediction"), as in [7]. Additional classes of problems rely on algorithms and techniques that apply to raw data at an earlier "pre-processing" stage. Different data processing may be employed if evaluation methodology is essential, [1].

2. STATISTICAL PARADIGMS

The adoption of the Guide to the Expression of Uncertainty in Measurement (GUM), [8] has led to an increasing need to include uncertainty statements in measurement results. The laboratory accreditation based on standards such as ISO 17025, [9] has accelerated the process. Recognizing that uncertainty statements are required for effective decision making, different laboratories, from national metrology institutes to commercial test laboratories, put considerable effort into analysis of measurement uncertainty using the GUM methods, [8, 10]. The methods for uncertainty assessment conducted in the TIC applications include the frequentist, Bayesian, and fiducial statistical paradigms, [11].

The first statistical paradigm, in which uncertainty can be probabilistically evaluated, is frequentist (based on statistical theory, referred as "classical" or "conventional"). Due to the nature of uncertainty in TIC, these methods must be adapted to obtain frequentist uncertainty intervals under realistic conditions. In most practical TIC settings, uncertainty intervals must consider both the uncertainty in quantities estimated using data and the uncertainty in quantities based on expert knowledge, i.e. data fusion. To obtain an uncertainty interval, the measurands that are not observed must typically be treated as random variables with probability distributions for their values, whereas measurands whose values can be estimated using statistical data are treated as unknown constants. The traditional frequentist procedures must be modified to attain the specified confidence level after averaging over the potential quantities values assessed by expert judgment [11].

The second paradigm is called the Bayesian approach, [11], named after the fundamental theorem, which was proved by the Reverend Thomas Bayes in the mid-1700s. The analyst's knowledge about the measurands is modeled as a set of random variables following a probability distribution in the joint parameter space. The theorem allows the probability distributions to be updated based on the observed data and the inter-relationships of the parameters defined by the function or equivalent statistical models. Then, one obtains a probability distribution describing one's knowledge of measurand given the observed data.

The third statistical paradigm is the fiducial approach, developed by R. A. Fisher in the 1930s [11]. The probability distribution, (fiducial distribution) for measurand conditional on the data, is obtained from the interrelationship of measurand and the input value described by the function and the distributional assumptions on the data used to estimate.

3. DATA FUSION, DECISION-MAKING, AND RISK ANALYSIS IN TIC

The ultimate idea underlying data fusion is to obtain greater quality information for a specific purpose by exploiting the synergy of data gathered from different sources. Data fusion is the process of combining data or information to estimate or predict entity states, [12]. Applied in many decision-making domains, as TIC, it encompasses classification and pattern recognition used to support decisions. In TIC it is crucial not only to fuse data obtained from multiple sensors, but also to assess threats and risk. Data fusion increases robustness and reliability and reduces the vulnerability of the system supporting the decision, allowing decision-making even in absence of malfunction of some sources of information. It provides a better and larger coverage of space and time, reduces ambiguity, as better information provides better discrimination between available hypotheses. Data fusion is based on experimental data output by sensing devices and on information obtained by other means (e.g. the user as a data source for a priori knowledge, experience, and model application). Fusion requires all data to have the same representation (e.g., numeric values in the same units, relative values). If data are heterogeneous, data alignment or data registration is imposed [12]. Measurements, as sensor output, form a signal more or less affected by noise whose reliability has to be verified (e.g. sensor malfunction, express corruption of sensors' measured quantity, e.g. jamming). Data filtering and data validation are two important tasks in a data fusion. Data fusion involves addressing: data from sources with different quality levels (e.g. different accuracy), non-independent data, too much information, leading to computational problems, and need to change the context of the observation (e.g. from time to frequency domain) or to extract features or attributes [12], as issues of data-processing type.

As an illustration for application of data science in the TIC sector, one of the most relevant TIC decision-making and risk-introducing issues will be further demonstrated - determination of the re-calibration period of the TIC measurement equipment.

4. APPLICATION OF DATA FUSION FOR DETERMINING THE TIC INSTRUMENT RE-CALIBRATION PERIOD

Determining re-calibration intervals is a problem of TIC sector entities using calibrated instruments in their activities. Most of the test equipment in today's inventories are multiparameter items or consist of individual single-parameter items. An item is stated to be out-of-tolerance if a single instrument parameter or item in a set, is found out of specifications. This is costly and introduces risks [13, 14]. Most of the published methods are of statistical nature and can correctly be applied only to large inventories of instruments, [15]. Due to the various performance of individual instruments and their different operation conditions, individual product reliability is difficult to predict. Longer calibration intervals have a higher consequence cost associated with a given standard, as more calibrations have been performed before it is re-calibrated and found to be in- or out-of-tolerance. Consequence costs may include a reverse traceability analysis to identify the items that have been calibrated by the standard, an investigation of the impact on their performance given the magnitude of the standard’s out-of-tolerance, customer notification, suspension of accreditation, product recall and intangible factors like the TIC entity’s reputation. Here the emphasis is on determining the recalibration period of measuring instruments used by the TIC entities. The methodology for determining the recalibration interval will be validated in a case study on experimental calibration and check data of an electrical measuring instrument, by fusion of data from diverse sources (both experimental and a priori knowledge, experience, model application). Most of the standards according to which the TIC entities are accredited/certified require to have available, suitable and adequate facilities and equipment to permit all TIC activities to be carried out in a competent and safe manner, with the responsibility lying solely on the TIC entity. One of the most significant decisions regarding the calibration is “When and how often to do it?” Many factors influence the time interval between calibrations and should be taken into account by the TIC entity. The most important factors are: uncertainty of measurement required or declared by the laboratory, risk of a measuring instrument exceeding the limits of the maximum permissible error when in use, cost of necessary correction measures when it is found that the instrument was not appropriate over a long period of time, type of instrument, tendency to wear and drift, manufacturer’s recommendation, extent and severity of use, environmental conditions (climatic conditions, vibration, ionizing radiation, etc.), trend data obtained from previous calibration records, recorded history of maintenance and servicing, frequency of cross-checking against other reference standards/measuring devices, frequency and quality of intermediate checks in meantime, transportation arrangements and risk, and degree to which the personnel are trained [15]. The ILAC-G24 specifies the following methods, [15]: automatic adjustment or “staircase” (calendar-time), control chart (calendar-time), “in-use” time, “in service” checking, or “black-box” testing, and other statistical approaches. The statistical methods, i.e. by deploying data science, of an individual instrument or instrument type are of interest, especially combined with adequate software tools. According to Agilent Technologies®, prior to the introduction of a new product, [17] the responsible personnel set the initial recommended calibration interval. The reliable data is from at least three areas: data from similar products, data for the individual components used in the instrument, data on any subassemblies leveraged from existing mature products. The typical operating conditions and the results of the environmental testing performed on product prototypes are also considered [18]. Several methods for determining the calibration intervals are published, [13], [14]. Some models are based on the assumption that the calibration condition of the instrument can be traced by monitoring the drift of an observable parameter, [13]. The calibration intervals can be presented according to analysis by parameter variables data, analysis by parameter attributes data, instrument attributes data and by class instrument attributes data. Other method, such as an extension by providing a maximum likelihood estimation for the analysis of data characterized by unknown failure times, are given in [13], where the estimation method is using the exponential reliability function. An approach with a review of the instrument’s calibration history are presented in [14], calibration record indicates a history of remaining in tolerance, as it might be expected that the instrument might have a higher likelihood of remaining in tolerance, as a result an algorithm that has been developed calculates calibration intervals based on the condition received at calibration along with a historical weighting. A method from variables data are presented for determining calibration intervals for parameters whose value demonstrate time-drift with constant statistical variance. The method utilizes variables data in the analysis of the time-dependence of deviations between as-left and as-found values from calibration. The deviations are from the difference between a parameter's as-found value at a given calibration and as-left value prior to calibration [14].

4.1. A model for determination of a re-calibration interval

Table 1: Values of parameters as multipliers

Parameter		Value
X	"In Tolerance"	1
	"Out of Tolerance" < 1x the tolerance band	0.8
	"Out of Tolerance" > 1x the tolerance band,< 2x the tolerance band	0.6
	"Out of Tolerance" > 2x the tolerance band,< 4x the tolerance band	0.4
	"Out of Tolerance" > 4x the tolerance band,< 4x the tolerance band	0.3
Y=ΣYi	Y_1 number of in-service checks between calibrations	
	1 time	0.1
	< 5 times	0.3
	< 10 times	0.4
	> 10 times	0.5
	Y_2 measured value	
	no difference (<3%)	0.5
	difference < 20%	0.4
	difference > 20%	0.1
Z=ΣZi	Z_1 Frequency of usage	
	dayly	0.1
	montly	0.5
	yearly	0.7
	Z_2 Habit of usage	
	used with caution in laboratory conditions	0.3
	used with caution in tendency to wear and drift	0.2
	use without special attention in terms of events	0.1
U=ΣUi	U_1 Cost of calibration	
	Small	0
	Medium	0.3
	Large	0.5
	U_2 Cost of necessary correction measurement	
	< 0.5 x cost of calibration	0.5
	< 1 x cost of calibration	0.1
	> 1 x cost of calibration	0
V	The operator is trained to handle the instrument and knows the measured items	1
	The operator is trained to handle the instrument, but imperfectly acquainted with the measured items	0.5
W	Service performed between previous and last calibration	0.5
	No service performed between previous and last calibration	1

Based on the previous discussions and survey, the following innovative data fusion model for determination of the re-calibration period is proposed

$$NI = ECI \cdot [C_1 \cdot X + C_2 \cdot X + C_3 \cdot X + C_4 \cdot X + IC \cdot Y + \\ + CFU \cdot Z + CO \cdot U + OFH \cdot V + MS \cdot W] \qquad (1)$$

where:

NI **-** New Interval
ECI **-** Established Calibration Interval
C_1 **-** Most recent calibration (0.2, modification of the simplified method)
C_2 **-** Most Previous Calibration (0.1, modification of the simplified method)
C_3 **-** Previous Calibration (0.08, modification of the simplified method)
C_4 **-** Predicted Value of Next Calibration (0.06, newly introduced parameter)
IC **-** In-service check between calibrations (0.1, newly introduced parameter)
CFU **-** Condition and Frequency of Use (0.2, newly introduced parameter)
CO **-** Costs (0.1, newly introduced parameter)
OFH **-** Operator factor and habit (0.08, newly introduced parameter)
MS **-** Maintain and service (0.08, newly introduced parameter)

ECI can be specified depending on the experience with the stability of similar instruments, experience and recommendations. It will be the longest possible re-calibration period, leading to a conclusion that the method is more rigorous in comparison to the "simplified method". The parameters as multipliers are given in Table 1.

4.2. Method validation

For proper application of the proposed model, a data base containing the historical data of previous calibrations must be created by the TIC entity. This is recommended to be applied after at least two performed calibrations. As case study for validation of the proposed method, a real data base with the calibration history of a digital multimeter used during process of inspections in a TIC body is adopted. The calibration values fluctuation should be observed in as many as possible measurement points, especially in points where changes are detected. Reasonable value for X is the smallest value that is obtained from all points. The expected value of the next calibration can be derived with a sophisticated algorithm, but for some TIC entities it is a complicated methodology. The proposed algorithm, using the least squares method, is a readily available and simplified tool. The in-service checks with another instrument must be performed in points where uncertainty of calibration is available for the both instruments. Quality management determines the extent of factors and habits of the staff, while the instrument operator specifies the frequency and conditions of use. Depending on the available history data and tracking behaviour of the instrument, the coefficients proposed in the algorithm can be modified and customized for each instrument or group of instruments.

As a case study implementing the proposed algorithm, the calibration periods of a METREL Eutotest XE MI 3102 tester is calculated. According to Metrel®, [17] regular 6-months or 1-year calibration of all measurement functions of the instrument is recommended. In Tables 2 and 3 the calibration history for the instrument in a single point of the current and voltage measurement ranges are given, respectively. Reference calibration value for the current is 10 A, while the voltage is 400 V. The measurement uncertainty is divided by a coverage factor as the calibration is performed in different laboratories at factor of coverage k=1.65 (for rectangular distribution at probability of 95%) and k=2 (for normal distribution at probability of 95%).

Table 2: The calibration history in a single point of current measurement range of METREL Eurotest XE MI 3102

t [month]	I [A]	Uncertainty [mA]
0.00	9.98	0.020
15.00	10.00	4.121
36.80	10.01	14.545
52.60	9.98	14.545
72.90	10.00	0.015

Table 3: The calibration history in a single point of voltage measurement range of METREL Eurotest XE MI 3102

t [month]	U [V]	Uncertainy [V]
0.00	399.00	0.42
21.80	401.00	0.36
37.66	401.00	0.36
58.02	400.00	0.50

The trend line for both areas is calculated without the inclusion of the last calibration value, which is used for verification of the model and it is:

$$U = -0.0024 \cdot t^2 + 0.1448 \cdot t + 399$$
$$R^2 = 1 \tag{2}$$

$$I = -7 \cdot 10^{-7} t^3 + 10^{-5} t^2 + 0.0013t + 9.98$$
$$R^2 = 1 \tag{3}$$

In Figures 1 and 2 the values of the function modeled including the last calibration, are shown. So the expected values are 9.85 A for current measuring range and 399.32 V for the voltage measurement range, and are in tolerance.

In the Table 4 the experimental results of in-service check measurements with two instruments, are displayed. The results are in limits of errors.

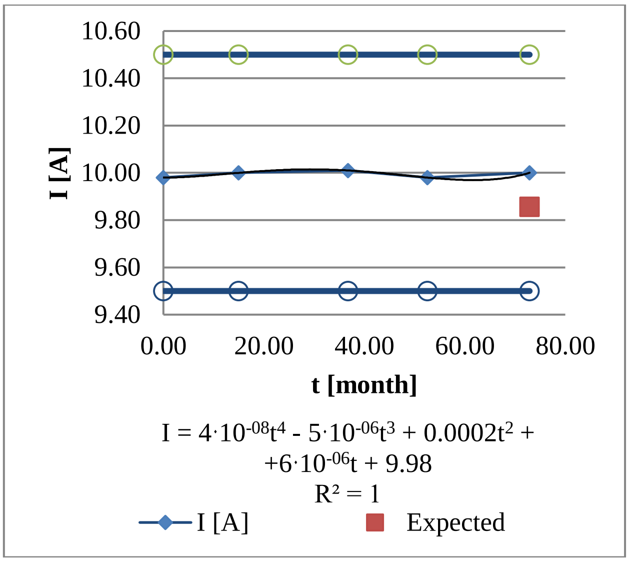

Figure 1: The calibration history and expected value in a single point of current measurement range

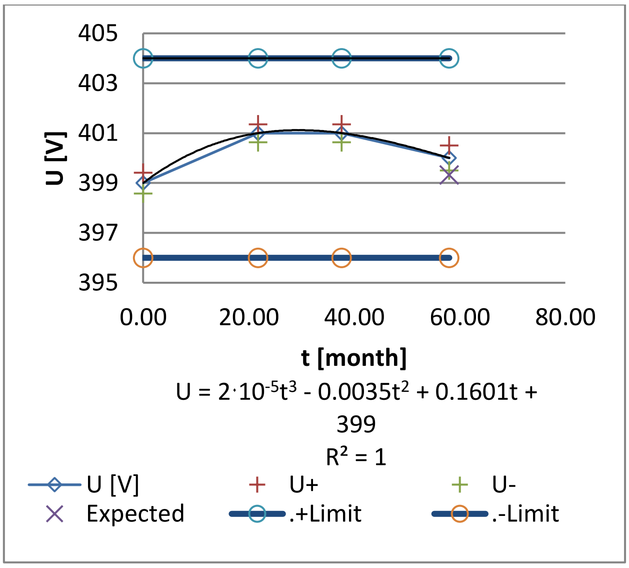

Figure 2: The calibration history and expected value in a single point of voltage measurement range

Table 4: Experimental values of the last in-service check with other instrument

Instument	I [A]	U [V]
Metrel MI 3102	3.7	224
Metrel MI 833	3.7	223

Other values for the parameters in the algorithm are as follows:

ECI = 24 months

$$\begin{aligned} NI &= ECI \cdot [C_1 \cdot X + C_2 \cdot X + C_3 \cdot X + C_4 \cdot X + \\ &+ IC \cdot Y + CFU \cdot Z + CO \cdot U + OFH \cdot V + \\ &+ MS \cdot W] = 24 \cdot [0.2 \cdot 1 + 0.1 \cdot 1 + 0.08 \cdot 1 + \\ &+ 0.06 \cdot 1 + 0.1 \cdot 0.8 + 0.2 \cdot 0.3 + 0.1 \cdot 0 + \\ &+ 0.08 \cdot 1 + 0.08 \cdot 1] = 24 \cdot 0.74 = 17.76 \end{aligned} \tag{4}$$

Table 5 .: Multipliers used in the case study

Parameter		Value
X	"In Tolerance"	1
Y=ΣYi	Y_1 number of check between calibration	
	< 5 times	0.3
	Y_2 measured value	
	no difference (3%)	0.5
Z=ΣZi	Z_1 Frequency of usage	
	dayly	0.1
	Z_2 Habit of usage	
	Used with caution in tendency to wear and drift	0.2
U=ΣUi	U_1 Coft of calibration	
	Small	0
	U_2 Cost of necessary correction measurement	
	> 1 x cost of calibration	0
V	The operator is trained to handle the instrument and knows the measured items	1
W	No service performed between previous and last calibration	1

The last calibration is not used in the prediction of the next value and is used as a validation point of the algorithm. The real calibration period (between the last two calibrations) is 20 months, while the predicted re-calibration period by the proposed algorithm is 18 months. The values obtained with the last calibration validate the method, and shorter value of the re-calibration interval is obtained which is on the safe side, which can be accepted as applicable.

8. Conclusion

The presented algorithm for predicting the period of re-calibration based on data fusion concept is simple, containing a lot of data on factors influencing the stability of the instrument derived from diverse sources. It is easily applicable in every day routine of various TIC entities. This algorithm can reduce the management risk of the occurrence of errors. The experimental values applied in the case study validate and confirm the effectiveness of the proposed algorithm. Another advantage of this universal model is that it allows the variation of the coefficients and enables specialization for a group of instruments. The data fusion approach is highly adaptable for various decision making applications in the TIC sector.

5. REFERENCES

[1] K. Cunha, R. Santos. The Reliability of Data from Metrology 4.0. International Journal on Data Science and Technology. Vol. 6, No. 4, 2020, pp. 66-69 https://doi.org/10.11648/j.ijdst.20200604.11

[2] B. J. Dorr, C.S. Greenberg, C.S., Fontana, P. et al. A new data science research program: evaluation, metrology, standards, and community outreach. Int J Data Sci Anal 1, 177–197, 2016 https://doi.org/10.1007/s41060-016-0016-z

[3] F. Pavese "An Introduction to Data Modelling Principles in Metrology and Testing", ed. F. Pavese, A. Forbes "Data Modeling for Metrology and Testing in Measurement Science", Birkhauser Boston, Springer Science+Business Media, LLC 2009, pp. 1-30

[4] F. R., Haas, L., Hernández, M., Miller, R.J., Popa, L., Velegrakis, Y.: Conceptual Modeling: Foundations and Applications. Springer, NY, 2009

[5] Sleeman, J., Finin, T., Joshi, A.: Entity type recognition for heterogeneous semantic graphs. In: 2013 AAAI Fall Symposium Series, 2013

[6] Jeevan, M.: Fundamental methods of data science: Classification, regression and similarity matching. http://www.kdnuggets.com/2015/01/fundamental-methods-data-science-classification-regression-similarity-matching.html, 2015

[7] Bengio, Y., Goodfellow, I.J., Courville A.: Deep learning. http://www.iro.umontreal.ca/bengioy/dlbook, 2015

[8] "Evaluation of measurement data — Guide to the expression of uncertainty in measurement", JCGM 100:2008 GUM

[9] EN ISO/IEC 17025"General requirements for the competence of testing and calibration laboratories", Cenelec, Brussels, 2017

[10] "Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method", JCGM 101:2008

[11] W. F. Guthrie, H. Liu, A. L. Rukhin, B. Toman, J. C. M. Wang, N. Zhang "Three Statistical Paradigms for the Assessment and Interpretation of Measurement Uncertainty", ed. F. Pavese, A. Forbes "Data Modeling for Metrology and Testing in Measurement Science", Birkhauser Boston, Springer, LLC 2009, pp. 71-116

[12] P. S. Girao, O. Postolache, J. M. D. Pereira "Data Fusion, Decision-Making, and Risk Analysis: Mathematical Tools and Techniques", ed. F. Pavese, A. Forbes "Data Modeling for Metrology and Testing in Measurement Science", Birkhauser Boston, Springer, LLC 2009, pp. 205-254

[13] A. Krleski, M. Cundeva-Blajer "Analysis and Contribution to Methods for Determination of Optimal Recalibration Intervals", Conf. Proc. of MMA'16 26th Symposium Metrology and Metrology Assurance 2016, 7-11 September 2016, Sozopol, Bulgaria, pp. 493-499

[14] H. Castrup "Calibration Intervals from Variables Data", NCSLI Work.&Symp. Wasington, 2005.

[15] ILAC-G24:2007 Guidelines for the determination of calibration intervals of measuring instruments.

[16] Setting and Adjusting Instrument Calibration Intervals – App. Note, Agilent Technologies, 2013.

[17] Eurocheck CS 2099 User manual, Metrel, 2013.

[18] Allen Bare et al. Simplified Calibration Interval Analysis, NCSLI Workshop & Symposium, 2006.