2025 IMEKO Joint Conference
TC8 - TC11 - TC25
Torino, Italy, September 14-17, 2025

Risk assessment for calibration: The non-linearized case

Dubravka Božić[1], Biserka Runje[1], Andrej Razumić[2]

[1] *Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lučića 5, Zagreb, Croatia*
dubravka.bozic@fsb.unizg.hr, biserka.runje@fsb.unizg.hr
[2] *Department of Polytechnics, Dr. Franjo Tuđman Defense and Security University, Ilica 256b, Zagreb, Croatia, andrej.razumic@sois-ft.hr*

***Abstract* – This study presents a novel mathematical model designed to evaluate risks in calibration processes. A probabilistic model was described for estimating the global producers' and consumers' risk with a non-linearized tolerance interval. The measurement uncertainty required to assess the risk for the explanatory variable at a given value of the response variable was calculated from the regression analysis outcome. The model was tested on the example from industrial practice to evaluate the risks associated with calibrating the roundness measurement device within a moderate scale of -3 µm to 3 µm. The results indicate that global consumer and producer risks are lower in the non-linear model than in the linear model. The calibration risk assessment model was evaluated using confusion matrix-based metrics.**

I. INTRODUCTION

In the process of product quality assessment, it is verified whether the product meets the prescribed standards. This process typically relies on measuring one or more key characteristics of the product. Examples include the quality assessment of water, food, fuels, industrial products, pharmaceuticals, and medical treatments [1–6]. An essential aspect of this process is evaluating the risk of wrongly rejecting a product that meets compliance standards, known as the global producer's risk (R_P). Conversely, the global consumer's risk (R_C), refers to the risk of accepting and using a product that fails to meet specified requirements. A widely recognized method for determining these risks is outlined in *Evaluation of measurement data – The role of measurement uncertainty in conformity assessment*, JCGM 106:2012 [7]. This method integrates knowledge of the true value of the characteristic of interest and its possible variation, with data from the measurement process. The true value of the characteristic property of interest is treated as an unknown parameter, modeled as a random variable *Y* whose probability density function is referred to as the prior. The prior can be any distribution, depending on the values that the measured characteristic can take [8]. Data from a future, independent measurement process are considered as a random variable Y_m and was modelled via a likelihood function. Knowledge about the prior and the likelihood function is combined using Bayes' theorem, which enables risk calculation within a given domain in a probabilistic manner [9,10]. However, the prescribed method for assessing the global producers' and consumers' risk can determine the risk for only a single point. Recently, it has been shown that the method can be extended two-dimensionally and three-dimensionally to data functionally related through a regression line [11]. The risk assessment model for calibration can be easily derived from the regression model [12]. In this study, a risk assessment procedure is described as a calibration model with a non-linearized tolerance interval. Risk modeling is demonstrated through the calibration of roundness measurement devices over a moderate range of −3 µm to 3 µm.

II. INPUT PARAMETERS OF THE MODEL

A. Transformation of the argument of the prior

Prior modeling requires information about the value of the best estimate y_0 of the characteristic of interest for a given item and the measurement uncertainty u_0 associated with the best estimate. To define the non-linearized (NL) model for risk assessment in calibration, it is assumed that the measurement data are functionally related through a calibration line whose equation is of the form $y = ax + b$, where $a \neq 0$ is the slope and b is the y-intercept of calibration line. The values of calibration line $y_i, i = 1,2,...,n$ in the point of the moderate scale $x_i, i = 1,2,...,n$ are calculated from the equation:

$$y_i = a\,x_i + b, i = 1,2,...,n. \quad (1)$$

For a given value of the response variable $y_{r_i}, i = 1,2,...,n$ the values of the explanatory variable are calculated from:

$$y_{0_i} \equiv x_{e_i} = \frac{y_{r_i} - b}{a}, i = 1,2,...,n. \quad (2)$$

These values in the model represent the best estimates $y_{0_i}, i = 1,2,...,n.$ To each value of the explanatory variable $x_{e_i}, i = 1,2,...,n$ is assigned a measurement uncertainty $u_{0_i}, i = 1,2,...,n$ calculated according to the law of propagation of error [13]. These measurement uncertainties were determined using data obtained in the calibration line fitting process and were calculated according to the equation:

$$u_0(x_{e_i}) = u_{0_i} = \sqrt{\frac{s_y^2 + u^2(b)}{a^2} + \frac{(y_i - b)^2 u^2(a)}{a^4}}. \quad (3)$$

The procedures for calculating the values $u(a)$ and $u(b)$ are omitted for simplicity and can be found in [11]. Equation (3) is valid for an equidistant moderate scale symmetric about zero. The prior is defined for each value $x_{e_i}, i = 1,2,...,n$ such that: $x_{e_i} \sim N(y_0, u_0^2), i = 1,2,\ldots,n.$ This completed the transformation of the prior argument of the pointwise risk assessment method outlined in [7]. The transformation was conducted by taking a continuous function as the argument of the prior. To calculate the risks, the discretization of the function described by Eq. (2) is performed. This discretization enables the representation of risk curves along the moderate scale axis.

B. Domain discretization

In addition to the discretization of the input function of the prior argument, it is necessary to define and perform the discretization of the domain over which the calculation for R_C and R_P are carried out. The input parameters of the model that define this domain are the tolerance interval $[T_L,T_U]$ and the acceptance interval $[A_L,A_U]$. The assessment of the global producer's and consumer's risk is conducted based on a decision rule with a guard band *w* placed between tolerance and acceptance intervals. In the context of the relationships between the tolerance and acceptance intervals, three models can be considered: the model of minimization of global producer's risk, where $[T_L,T_U] \subset [A_L,A_U]$, the shared risk model, where holds that $[T_L,T_U] = [A_L,A_U]$ and the model of minimization of global consumer's risk, where $[T_L,T_U] \supset [A_L,A_U]$. The mentioned models can be connected by joining their guard bands, enabling the risk curves to be displayed along the guard band axis [4,8].

In the linearized model (L), the tolerance and acceptance lines are parallel to and symmetrically positioned around the line y = x. The similarly can be applied to the tolerance and acceptance curves in the non-linearized model (NL). This setup assumes that, in a well-executed calibration, measured values closely match the values of the moderate scale points $x_i, i = 1,2,...,n.$ Therefore, the measurement uncertainty and risk were observed relative to y = x line. The motivation for non-linearized model derives from the shape of confidence and predictive intervals traditionally placed around the regression line. Such a definition of tolerance and acceptance curves allows for greater deviations of measured values at the ends of the moderate scale due to systematic and systemic measurement errors.

The upper and lower tolerance curves at points $x_i, i = 1,2,...,n$ are calculated from:

$$T_{U_i} = x_i + b_U, i = 1,2,...,n \quad (4)$$

and

$$T_{L_i} = x_i - b_L, i = 1,2,...,n. \quad (5)$$

Assuming that the tolerance curves are positioned symmetrically around the line y = x, it holds that $b_U = b_L > 0.$ To introduce non-linearity, it is set that $b_U = b_L = 3u_0(x_{e_i}) = 3u_{0_i}, i = 1,2,...,n.$ In the linearized model used for comparison, it is assumed that $b_U = b_L = 3\min(u_{0_i}), i = 1,2,...,n.$ For the model of minimization of global producer's risk holds that $\Delta T = 0.9\Delta A$ where is $\Delta A = A_U - A_L$ and $\Delta T = T_U - T_L.$ For the model of minimization of global consumer's risk holds that $\Delta A = 0.9\Delta T.$ It can be easily shown that the total width of the guard band in the NL model is $w_u = 0.1\Delta T = 0.6u_{0_i}, i = 1,2,...,n.$

The discretization of the acceptance interval is performed by introducing a multiplicative factor $r_k \in [-1,1]$, where $r_k = -1.1 + 0.1k, k = 1,2,...,m,\ m = 21.$ The lower and upper acceptance curves are calculated according to the following equations:

$$A_{L_{i,k}} = T_{L_i} + r_k \frac{w_U}{2}, i = 1,2,...,n, k = 1,2,...,m \quad (6)$$

and

$$A_{U_{i,k}} = T_{U_i} - r_k \frac{w_U}{2}, i = 1,2,...,n, k = 1,2,...,m. \quad (7)$$

The risk curves along the guard band axis are plotted for all guard band widths $w_k = (r_k w_U)/2, k = 1,2,...,m.$ In this way, the acceptance interval limits change from the model of minimization of global producer's risk to the model of minimization of global consumer's risk. It is important to notice that Equations (4−7) define the tolerance and

acceptance intervals for the points $y_i, i = 1,2,...,n$ on the regression line. In the calibration model, it is necessary to define these intervals for the explanatory variable values from Eq. (2), as illustrated in Fig. 1.

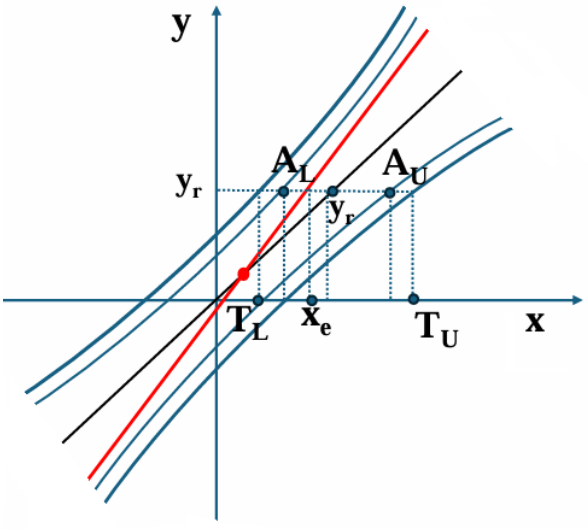

Fig. 1. Geometric representation of the non-linearized model for risk assessment in calibration; Model of minimization of global consumer's risk.

It can be easily shown that for a given response variable value y_{r_i} the tolerance interval of the explanatory variable x_{e_i} takes the form $\left[y_{r_i} - b_U, y_{r_i} + b_L\right]$, while the acceptance interval is given by $\left[y_{r_i} - b_U + w_k, y_{r_i} + b_L - w_k\right]$, where $i = 1,2,...,n$ and $k = 1,2,...,m$.

To model the likelihood function, it is essential to measure or assume the uncertainty u_m of a future measurement process, which should be independent of the process used to obtain the prior data. In this study, it is assumed that $u_{m_i} = u_{0_i}/2, i = 1,2,...,n$. The likelihood function is modelled under the assumption that the data from the measurement process are normally distributed.

III. RISK CALCULATION

Assume that $T = \left[T_L, T_U\right]$, $T^C = \left(-\infty, T_L\right] \cup \left[T_U, \infty\right)$, and similarly, $A = \left[A_L, A_U\right]$, $A^C = \left(-\infty, A_L\right] \cup \left[A_U, \infty\right)$. If the values taken by the random variable Y_m fall outside the acceptance interval, while the values of the random variable Y fall within the tolerance interval, then the global producer's risk can be calculated according to the equation:

$$R_P = \int_T \int_{A^C} g_0(\eta) h(\eta_m \mid \eta) d\eta_m d\eta. \tag{8}$$

Conversely, if the values of the random variable Y_m fall within the acceptance interval, while the values of the random variable Y are outside the tolerance interval, the global consumer's risk is calculated as:

$$R_C = \int_{T^C} \int_A g_0(\eta) h(\eta_m \mid \eta) d\eta_m d\eta. \tag{9}$$

In Equations (8) and (9), $g_0(\eta)$ denotes the prior distribution and $h(\eta_m|\eta)$ denotes the likelihood function of the true value η given a measured quantity value η_m [7]. The procedure for reducing double integral to a single integral, as well as the final risk calculation formulas after prior transformation and domain discretization, are detailed in [7,12]. These details are omitted here for clarity and brevity.

IV. RESULTS AND DISCUSIONS

A. Results of calibration line fitting

The global producer's and consumer's risks in the NL model were evaluated during the calibration of an inductive contact probe device for roundness measurement (Mahr MMQ3). Measurements were performed in the Laboratory for Precise Measurement of Length at the University of Zagreb in Croatia. At each point of the moderate scale $x_i, i = 1,2,...,n$, $n = 13$ three measurements were conducted [14]. The calibration line was determined based on the mean value of these measurements. Due to the autocorrelation of residuals detected by the Durbin-Watson test, the calibration line was fitted according to the Hildreth-Lu method [15]. The estimated autocorrelation coefficient of the errors calculated by this method has a value ρ = 0.2. That coefficient value ensures an efficient estimation of the calibration line parameters and a consistent estimation of the standard errors. In this regime, the parameters of the calibration line were $a = 1.00839$ µm, and $b = -0.002784$ µm. The remaining parameters needed for the calculation of measurement uncertainties from Eq. (3) are: $\sigma_y = 0.0182$ µm, $u(a) = 0.003804$ µm, and $u(b) = 0.005375$ µm [12].

B. Risk curves and surfaces

Along the guard band axis, the global consumer's risk decreases, while the global producer's risk increases. Along the moderate scale, the risk curves are parabolas with a minimum at the intersection of the calibration line and the line y = x. This holds true for both the global producer's risk and the global consumer's risk. The minimum risk in the calibration of the inductive control probe is achieved at the point on the moderate scale $x_{\min} = 0.3317$ µm. These behaviors can be clearly observed by analyzing the risk surfaces. The risk surfaces for the NL model, along with highlighted curves of minimal risk, are illustrated in Fig. 2.

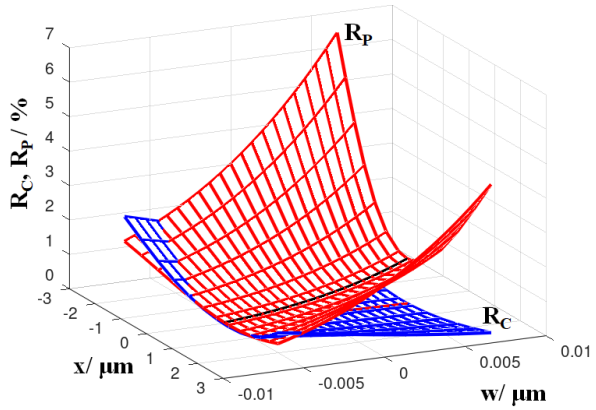

Fig. 2. Surfaces of the global producer's risk Rp and the global consumer's risk Rc.

The maximum risk of discarding a measurement as non-conforming is reached at the edges of the moderate scale. Greater risk values are observed at the negative end.

C. Comparison with a non-linearized model

The maximum and minimum values for R_C and R_P in the L and NL models are presented in Table 1 and Table 2. The coordinates of the points at which the minimum and maximum values are achieved are given in µm. The first coordinate of the listed points lies on the calibration scale axis, and the second coordinate lies on the guard band axis.

Table 1. Global consumer's risk

	(x,w) /µm, Min /%	(x,w) /µm, Max /%
L	$\left(x_{\min}, 5.65\cdot10^{-3}\right), 0.041$	$\left(-3, -5.65\cdot10^{-3}\right), 3.76$
NL	$\left(x_{\min}, 5.66\cdot10^{-3}\right), 0.04$	$\left(-3, -6.59\cdot10^{-3}\right), 1.9$

Table 2. Global producer's risk

	(x,w) /µm, Min /%	(x,w) /µm, Max /%
L	$\left(x_{\min}, -5.65\cdot10^{-3}\right), 0.187$	$\left(-3, 5.65\cdot10^{-3}\right), 9.12$
NL	$\left(x_{\min}, -5.66\cdot10^{-3}\right), 0.19$	$\left(-3, 6.59\cdot10^{-4}\right), 6.4$

The minimum measurement uncertainty associated with the best estimate $y_{0i}, i = 1,2,...,n$ occurs at the middle of the calibration scale for $i = 7$, measuring $u_{0\min} \approx 0.019$ µm. This is also the only value for which the b_U and b_L in the L and NL models coincide. Consequently, and due to the construction of the tolerance interval and acceptance interval, the global producer's risk and the consumer's risk of the L and NL model have equal values at points $(0, w_k), k = 1,2,...,m$. Across the remaining domain, the global producer's risk is consistently higher in the L model compared to the NL model. The same observation holds true for the global consumer's risk, as shown in Fig. 3.

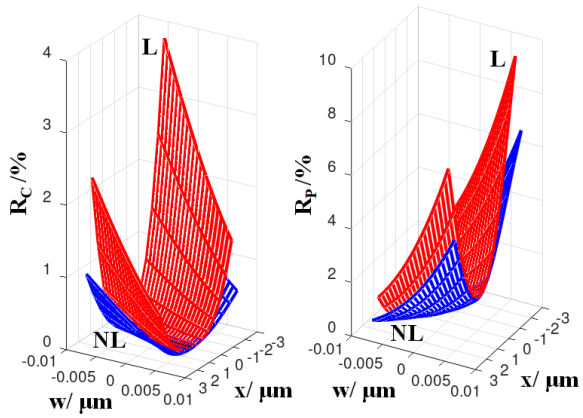

Fig. 3. Risk comparison of linearized and non-linearized models.

V. MODEL EVALUATION

The non-linearized model for risk assessment in calibration was evaluated by determining the probability of conformity p_C. That is the probability that an item of interest fulfills a specified requirement [7]. Considering that the prior and likelihood are normally distributed, the conformance probability for values of the explanatory variable from Eq. (2) is calculated using the following equation:

$$p_{C_i} = \frac{1}{u_{0_i}\sqrt{2\pi}} \int_{T_{L_i}}^{T_{U_i}} \exp\left[-\frac{1}{2}\left(\frac{\eta - x_{e_i}}{u_{0i}}\right)^2\right] d\eta, \qquad (10)$$

where $i = 1,2,...,n$ [12]. Since the NL model has lower values for R_C and R_P compared to the L model across the domain, it exhibits higher conformance probability values, as shown in Fig. 4.

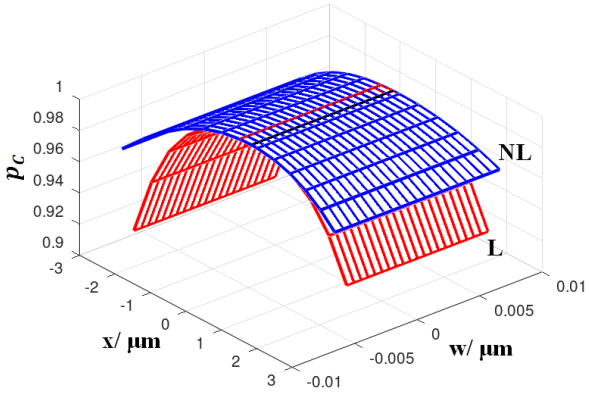

Fig. 4. Conformance probability; Comparison of linearized and non-linearized model.

The maximum and minimum values for the conformance probability for the L and NL models are given in Table 3.

Table 3. Conformance probability

	(x,w) /µm, Min /%	(x,w) /µm, Max /%
L	$(-3, w_k), 90.27^{*}$	$(x_{\min}, w_k), 99.72$
NL	$(-3, w_k), 95.97$	$(x_{\min}, w_k), 99.73^{**}$

$^{*}\ k = 1, 2, ..., m$, ** black highlighted line in Fig. 4

The conformance probabilities of the L and NL models are identical at $(0, w_k), k = 1, 2, ..., m$. These points are represented in Fig. 4 as a red highlighted line.

If the values taken by the random variable Y are within the tolerance interval and that the values taken by the random variable Y_m are simultaneously within the acceptance interval, the true positive probability (TP) can be defined. Analogously, if the values taken by the random variable Y are outside the tolerance interval and the values taken by Y_m are outside the acceptance interval, the true negative probability (TN) is defined. The probabilities TP and TN are calculated similarly to Eqs. (8-9), but they involve different areas of integration. It holds that p_C = TP+R_P and 1−p_C = TN+R_C [16]. These relationships enable the evaluation of the proposed model for risk assessment in calibration by using classification metrics derived from the confusion matrix. In each point on the moderate scale, it is possible to determine the probability of belonging to one of the classes TP, TN, R_C and R_P.

Among the many metrics for model evaluation, the F1 score was selected as suitable for imbalanced data [17,18]. The F1 score is the harmonic mean of the recall and precision metrics. In risk terms, the F1 score is calculated as follows:

$$F1 = \frac{p_C - R_P}{p_C - R_P + \dfrac{R_P + R_C}{2}}. \qquad (11)$$

The F1 score surfaces for the L and NL models are shown in Fig. 5.

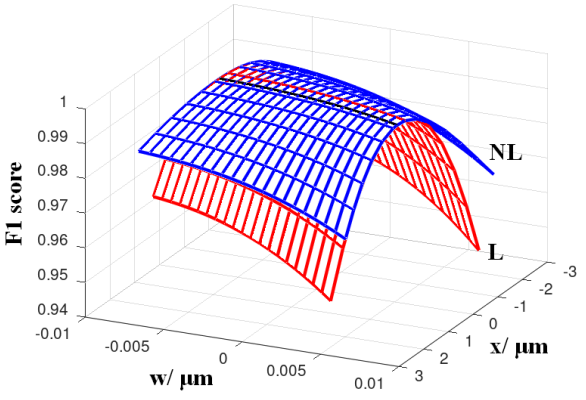

Fig. 5. F1 score surfaces.

The values for the F1 score of the L and NL models coincide at the points $(0, w_k), k = 1, 2, ..., m$, represented as the red curve in Fig. 5. The maximum and minimum values of the F1 score are given in Table 4.

Table 4. F1 score

	(x,w) /µm, Min /%	(x,w) /µm, Max /%
L	$(-3, 5.65 \cdot 10^{-3}), 94.03$	$(x_{\min}, -5.65 \cdot 10^{-3}), 99.83$
NL	$(-3, 6.59 \cdot 10^{-3}), 96.28$	$(x_{\min}, -5.66 \cdot 10^{-3}), 99.84$

The maximum values of the F1 score are achieved at the points $(x_{\min}, w_k), k = 1, 2, ..., m$. This curve, highlighted in black in Fig. 5, decreases along the guard band axis. Along the guard band axis, the F1 score curves also form parabolas with their maximum at the calibration scale point $x_{\min}$.

Although the L model exhibits higher peak values for global producer's and consumer's risks (see Fig. 3), the NL model outperforms it by achieving higher conformance probabilities and F1 scores. Furthermore, both the L and NL models show high values for p_C and the F1 score. This suggests a high-quality measurement process and a calibration line that closely aligns with y = x.

VI. CONCLUSION

The proposed calibration risk assessment method enables risk evaluation not only at calibration points but also across the entire measurement range. The method allows an assessment of the global producers' and consumers' risks for various guard band widths. Moreover, the method is suitable when measurement uncertainty or tolerance intervals are unknown, as these can be estimated directly from the calibration data. In addition, this straightforward method for risk assessment in calibration can be useful across different scientific disciplines.

REFERENCES

[1] L.P.Brandão, V.F.Silva, M.Bassi, E.C.de Oliveira, "Risk assessment in monitoring of water analysis of a Brazilian river", Molecules, vol.27, No.11, June 2020, 3628.

[2] F.R.Pennecchi, I.Kuselman, A.Di Rocco, D.B.Hibbert, A.A.Semenova, "Risks in a sausage conformity assessment due to measurement uncertainty, correlation and mass balance constraint", Food Control, vol.125, July 2021, 107949.

[3] E.C.de Oliveira, F.R.Lourenço, "Risk of false conformity assessment applied to automotive fuel analysis: A multiparameter approach", Chemosphere, vol.263, January 2021, 128265.

[4] D.Božić, M.Samardžija, M.Kurtela, Z.Keran,

B.Runje, “Risk Evaluation for Coating Thickness Conformity Assessment”, Materials, vol.16, No.2, January 2023, 758.

[5] L.Separovic, R.S.Simabukuro, A.R.Couto, M.L.G.Bertanha, F.R.S.Dias, A.Y.Sano, A.M.Caffaro, F.R.Lourenço, “Measurement Uncertainty and Conformity Assessment Applied to Drug and Medicine Analyses – A Review“, Crit. Rev. Anal. Chem., vol.53, No.1, Jun 2021, pp.123–138.

[6] L.Carvalheira, J.M.Lopes, P.F.de Aguiar, E.C.de Oliveira, “Compliance assessment when radioactive discharges are close to exemption levels in nuclear medicine facilities”, Radiat. Phys. Chem., vol.203, Part A, February 2023, 110636.

[7] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, “Evaluation of measurement data — The role of measurement uncertainty in conformity assessment”, Joint Committee for Guides in Metrology, JCGM 106:2012.

[8] D.Božić, B.Runje, “Selection of an Appropriate Prior Distribution in Risk Assessment,” Proc. of the 33rd International DAAAM Virtual Symposium “Intelligent Manufacturing & Automation”, 2022, pp. 26-27.

[9] I.Lira, “A Bayesian approach to the consumer's and producer's risks in measurement”, Metrologia, vol.36, No.5, October 1999, 397.

[10] E.T.Volodarsky, L.O.Kosheva, M.O.Klevtsova, “The Role Uncertainty of Measurements in the Formation of Acceptance Criteria”, Proc of 2019 XXIX International Scientific Symposium" Metrology and Metrology Assurance" (MMA), IEEE, 2019, pp. 1-4.

[11] D.Božić, B.Runje, A.Razumić, “Risk assessment for linear regression models in metrology”, Appl. Sci., vol.14, No.6, 2605.

[12] D.Božić, B.Runje, A.Razumić, “Risk assessment procedure for calibration”, Tehnički glasnik, vol.19, No.si1, June 2025.

[13] N.A.Heckert, J.J.Filliben, C.M.Croarkin, B.Hembree, W.F.Guthrie, P.Tobias, J.Prinz, “Handbook 151: Nist/sematech e-handbook of statistical methods”, 2012.

[14] F.Bednjanec, “Umjeravanje uređaja za mjerenje kružnosti”, (Peplowski rad), Zagreb, Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, (in Croatian), 2016.

[15] C.Hildreth, J.Lu, “Demand relations with autocorrelated disturbances”, Technical Bulletin. Michigan State University Agricultural Experiment Station, No.276, 1960, pp.16

[16] D.Božić, B.Runje, D.Lisjak, D.Kolar, “Metrics related to confusion matrix as tools for conformity assessment decisions”, Appl. Sci., vol.13, No.14, July 2023, 8187.

[17] L.A.Jeni, J.F.Cohn, F.De La Torre, “Facing imbalanced data-recommendations for the use of performance metrics,” Proc. of 2013 Humaine association conference on affective computing and intelligent interaction, IEEE, 2013, pp. 245-251

[18] D.Chicco, N.Tötsch, G.Jurman, “The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation”, BioData Min., vol.14, No.13, February 2021, pp.1-22.