applied sciences

MDPI

Article

Risk Assessment for Linear Regression Models in Metrology

Dubravka Božić *, Biserka Runje and Andrej Razumić

Department of Quality, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, 10000 Zagreb, Croatia; biserka.runje@fsb.unizg.hr (B.R.); andrej.razumic@fsb.unizg.hr (A.R.)
* Correspondence: dubravka.bozic@fsb.unizg.hr

Featured Application: This manuscript can be useful in decision making and quality control.

Abstract: The conformity assessment of products or a measured value with the given standards is carried out based on the global risk of producers and consumers' calculations. A product may conform to specifications but be falsely rejected as non-conforming. This is about the producer's risk. If a product does not meet the requirements but is falsely accepted as conforming, that poses a risk to the consumer. The conventional approach to risk assessment, which yields only a single numerical value for the global risk of producers and consumers, is naturally extended and utilized for assessing risk in measurement models with linear regression. The outcomes of the two-dimensional extension, along a moderate scale, are the parabolas with upwards openings. Risk surfaces were obtained through three-dimensional extension over the area limited by the moderate scale and guard band axes. Four models with different ranges of tolerance intervals were used to test this innovative method of risk assessment in linear regression. The corresponding standard measurement uncertainties were determined by applying a simplified measurement model with the use of comprehensive data on the measurement performance and by determining measurement uncertainty derived from consideration of the functional relationship obtained by linear regression analysis. Models that utilize information from linear regression analysis to determine measurement uncertainty are biased towards risks at the edges of the moderate scale. Testing the model's performances with metrics related to the confusion matrix, such as the F1 score, further substantiated this assertion. The diagnostic odds ratio has been proven to be extremely effective in identifying the curve along the guard band axis, along which the global risks of producers and consumers are at their lowest.

Keywords: regression; consumer's risk; producer's risk; tolerance interval; measurement uncertainty

check for updates

Citation: Božić, D.; Runje, B.; Razumić, A. Risk Assessment for Linear Regression Models in Metrology. *Appl. Sci.* **2024**, *14*, 2605. https://doi.org/10.3390/app14062605

Academic Editor: José Campos Matos

Received: 16 February 2024
Revised: 15 March 2024
Accepted: 18 March 2024
Published: 20 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Regression is the most used statistical method, applied for various purposes and in different scientific disciplines, including metrology. The basic purpose of regression is the prediction of a value that has not been measured, but which can be estimated based on the regression line or regression plane. So, for example, a regression can be used to predict the cutting temperature and surface roughness during the face milling process [1], or for predicting and monitoring vibrations on a lathe machine [2]. Regression can also be used to characterize various devices and sensors for atomic force microscope characterization [3] or magnetic field sensors characterization [4]. It is most often used in calibration, from calibrating pressure against ruby fluorescence shifts [5] to calibration that describes the pressure sensitivities of the optical fiber sensors [6]. Furthermore, regression is often used in interlaboratory comparison studies of calibration standards in different fields [7,8], or for the analysis of different physical constants [9,10].

The application of regression involves measuring the value of the response variable that corresponds to the given values of the explanatory variable [11]. There are numerous existing regression techniques that can be used to determine the relationship between these variables [12]. Among all of them, a straight-line relationship is the simplest one. This, the

most common way of connecting only one independent (explanatory) variable and the dependent (response) variable, is called univariate linear regression. The subject of this paper concerns the risk assessment for linear univariate regression models in metrology.

All recommendations and standards related to metrology are issued by the international metrology organization International Bureau of Weights and Measures (Bureau International des Poids et Mesures, BIPM). The basic procedure for risk assessment for an item of interest is given by the guide called Evaluation of measurement data—The role of measurement uncertainty in conformity assessment 106:2012 [13]. However, this method of risk assessment, well-known in metrology, has not been applied for regression so far. According to [14] (p. 15), and [15,16] guidelines called JCGM 107, for the application of the least squares method, the most famous method for determining the coefficients of the regression line, is still in preparation. It is not yet known whether these guidelines will include a risk assessment.

Risk assessment is carried out in the process of assessing the conformity of a product with specified requirements. Two types of risks may arise during conformity assessment. The producer's risk refers to situations where measurements or products meet the required specifications but are rejected as non-conforming. Consumer risk refers to situations when products or measurements are accepted as conformed but do not meet specifications. Whether a product conforms to the given specifications or not is determined based on measurements. The measured value of the item of interest must be within the given tolerance interval. The measurement uncertainty that occurs during measurement may lead to incorrect decisions regarding the acceptance of a non-compliant product or the rejection of a product that meets the specifications. This happens when the measurement is close to either the lower limit of the tolerance interval T_L or the upper limit of the tolerance interval T_U, and the measurement uncertainty associated with that measured value goes over the tolerance interval $[T_L, T_U]$ [17–21].

It is important to emphasize that the tolerance interval in the metrology domain is different from the regression tolerance interval in the statistical domain. The statistical tolerance regions, also known as simultaneous tolerance intervals, are constructed so that they, with a certain level of confidence, contain a specified proportion of the population in future sampling [22,23]. For reasons that will be elaborated upon further in the text, it is not feasible to apply these simultaneous tolerance intervals for the model of risk assessment in regression that is presented in this paper. The tolerance interval in the metrology domain is established by the manufacturer in its specifications for a particular measuring device, or this interval is determined by the applicable standards.

The risk assessment method outlined in this paper can be applied not only in metrology, but also in general in all regression models, regardless of the area in which the data originate. Therefore, this manuscript gives alternative values for the tolerance interval that can be used when the values for the tolerance interval provided by the manufacturer are unavailable.

To reduce the impact of measurement uncertainty on risk assessment, an acceptance interval $[A_L, A_U]$ is introduced into the procedure for evaluating the conformity of products with the prescribed specifications, in addition to the tolerance interval. The marks A_L and A_U represents the lower and upper limits of the acceptance interval, respectively. Acceptance and tolerance intervals can be in different mutual relations [24]. The limits of the tolerance interval and the acceptance interval depend on the properties of the measurand. These can be semi-open or semi-closed intervals [25]. Furthermore, they can be closed intervals, as in this case. If the tolerance interval is within the acceptance interval, it is a model of minimization of the producer's risk, as illustrated in Figure 1a. If the tolerance interval is outside the acceptance interval, it is a consumer risk minimization model (Figure 1c) [26]. A special situation when tolerance and acceptance interval overlap is when it is valid that $A_L = T_L$ and $A_U = T_U$. This is called shared risk (Figure 1b).

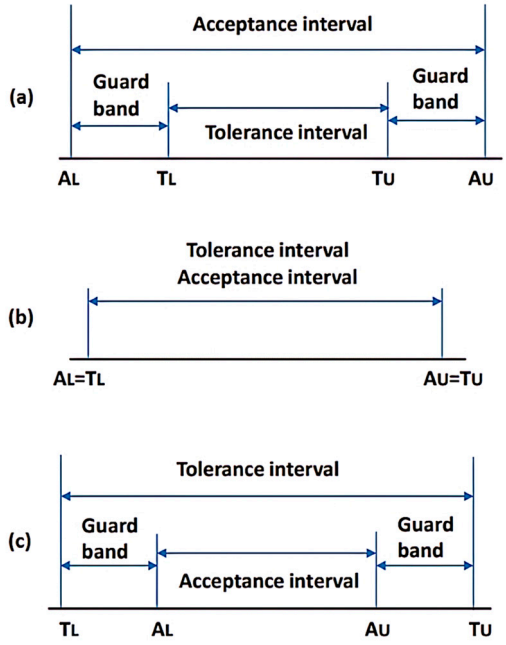


Figure 1. Interrelationships of tolerance interval and acceptance interval: (**a**) Minimizing producer risk; (**b**) Shared risk; (**c**) Minimizing consumer risk [26].

The guard band of length w, between the tolerance interval and the acceptance interval, ensures a reduction in the probability of making wrong decisions. In practice, it is commonly recommended to minimize the consumer's risk to enhance the quality of the delivered products. However, it is feasible to set the length of the guard band so that both the customer and the producer are satisfied.

When measuring, it is assumed that the object of interest has a measurable property Y with possible values η. It is natural to assume that there exists historical data about the item of interest, such as those found in manuals, scientific papers, data from previous measurements, etc. Also, one can talk about prior beliefs of the measurer based on experience regarding possible values that will be obtained during the measurement, or regarding the possible distribution of parameters that describe the measured data. Such data are treated as random variables. Knowledge about these data is given by prior distribution, that is, by the probability density function (PDF), designated with $g_0(\eta)$ [13]. Depending on the available data, the prior can be a non-parametric, one-parameter, or two-parameter distribution [27,28]. According to the principle of maximum entropy (PME), if data for two moments are available, a two-parameter distribution is used for the prior [29–31]. In this manuscript, the prior is a normal distribution whose parameters are the best estimate of a measurand y_0, with an associated standard measurement uncertainty u_0.

The data associated with the measurements Y_m are also treated as a random variable. The values that this random variable can take are denoted by η_m. Random variable Y_m is modeled via the likelihood function for normal distribution and denoted by $h(\eta_m|\eta)$ [13]. The formula for the likelihood function involves standard measurement uncertainty u_m of some future measurement. Global consumer's risk R_C and global producer's risk R_P are calculated probabilistically, using Bayes' theorem. This is carried out by combining information on the prior distribution and the likelihood function, that is, by combining information on the random variables Y and Y_m [13,19,32,33]. If the values for Y are outside the tolerance interval, and the values for Y_m within the acceptance interval, it is about the global consumer's risk R_C. If the values for Y are within the tolerance interval, and the values for Y_m are outside the acceptance interval, it is about the global producer's risk R_P.

The risk assessment method can be applied when product quality is evaluated based on only one measured quantity [13], for example, when determining the roundness deviation of the inner ring of the bearing [34] or when estimating the thickness of the epoxy coating

applied to water pipes [25]. This approach can also be used to evaluate conformity with the specifications for each distinct property of the item of interest [35]. It is also possible to assess the risks of multi-component models, where several factors can affect the quality of the product [36,37]. Examples include risk assessment in food quality control [38], drug quality control [39,40], air quality control [41], the chromatography process [42], pharmaceutical equivalence studies when comparing generic and reference drugs [43], and other related areas.

When the tolerance interval and acceptance interval are closed intervals, and in the case of risk evaluation for a single property of an item of interest, there exist two distinct models for the risk assessment: centered and non-centered models (Figure 2).

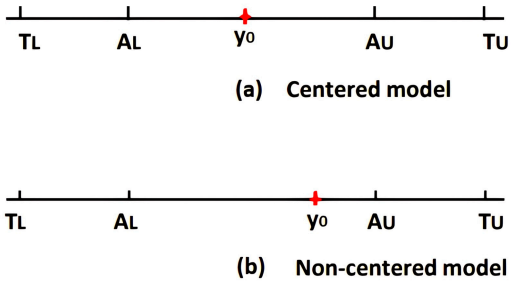


Figure 2. Minimization of consumer risk: (**a**) Centered model; (**b**) Non-centered model.

For the centered model, the best estimate of a measurand y_0 is exactly in the middle of the acceptance and tolerance interval, which is not the case with the non-centered model. This approach to risk assessment for a single measurable property, based on intervals of tolerance and acceptance, yields only one numerical value for the global risk of the producer, i.e., the consumer. It is evident that this method naturally can be extended to the two-dimensional case, where the outcomes are iso-risk curves, or to the three-dimensional case, where the results are risk surfaces. This extension of a simple approach for risk assessment allows for more reliable risk assessment in regression models in metrology.

2. Materials and Methods

2.1. Measurement Description

This article is focused on the development of a model for risk assessment in regression. Therefore, the measurement procedure utilized to collect the processed data is briefly described in the article.

The risk assessment was based on the information gathered during the calibration of the roundness measurement device, Mahr MMQ3. The calibration procedure consists of calibrating the inductive contact probe (dial indicator) on the universal length measuring machine (ULM) and determining the errors in the rotation of the measuring device spindle using the sphere standard [44]. All measurements were carried out in the Laboratory for Precise Measurement of Length of the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb. Key data for risk assessment were collected in the process of calibrating the inductive contact probe.

The risk assessment was carried out on a moderate scale from -30 µm to 30 µm. For each reference value $\mathbf{x} = (x_1, x_2, \ldots, x_{13})$ of the moderate scale, three measurements were performed: $\mathbf{y_i} = \left(y_{i,1}, y_{i,2}, \ldots, y_{i,13}\right)$, $i = 1, 2, 3$ (Table 1).

Table 1. Calibration of probes in the measuring range of ±30 µm [44].

Reference Value x/µm	1st Measurement y_1/µm	2nd Measurement y_2/µm	3rd Measurement y_3/µm
−30	−30.10	−30.02	−30.05
−25	−25.12	−25.05	−25.05
−20	−20.10	−20.05	−20.05

Table 1. *Cont.*

Reference Value x/µm	1st Measurement y_1/µm	2nd Measurement y_2/µm	3rd Measurement y_3/µm
−15	−15.10	−15.00	−15.05
−10	−10.10	−10.00	−10.05
−5	−5.10	−5.05	−5.06
0	0.00	0.00	0.00
5	4.98	4.98	5.00
10	9.99	10.00	10.02
15	14.99	15.00	15.00
20	20.00	20.00	20.05
25	25.00	25.02	25.00
30	30.00	30.05	30.00

Positive and negative reference values of the calibration scale indicate the direction of the probe movement during calibration.

2.2. Model Parameters

The input parameters needed for the calculation of the global consumer's risk R_C and global producer's risk R_P are as follows: the best estimate of a measurand y_0 with their associated standard uncertainty u_0, as well as the tolerance interval, acceptance interval, and measurement uncertainty u_m of a future inspection process.

2.2.1. Best Estimate of a Measurand

In assessing risk in regression models, the best estimates of a measurand y_0 can be obtained by calculating the values for the regression line y_f in the points of the reference scale **x**. In that case, the best estimate of the measured quantity is represented by the vector $\mathbf{y_0} = (y_f(x_1), y_f(x_2), \ldots, y_f(x_{13}))$, and the risk is estimated for each of the values of the regression line $y_{f_j} = y_f(x_j), j = 1, 2, \ldots, 13$. The coefficients $\hat{\beta}_0$ (intercept) and $\hat{\beta}_1$ (slope) of the regression line were obtained by the well-known least squares (LS) method from a simple linear regression model:

$$y_{i,j} = \beta_0 + \beta_1 x_j + \varepsilon_{i,j}, i = 1, 2, 3, j = 1, 2, \ldots, n_2, \quad (1)$$

where the intercept β_0 and slope β_1 are the model parameters and $n_2 = 13$. For simplicity, in the subsequent text, where appropriate, the designation $n_1 = 3$ is utilized. The values $\varepsilon_{i,j}, i = 1, 2, 3, j = 1, 2, \ldots, n_2$ are random errors [45,46]. Model parameters β_0 and β_1 were estimated for each pair of observations $\left(x_j, y_{i,j}\right), i = 1, 2, 3, j = 1, 2, \ldots, n_2$ [47] (pp. 76–78). The best estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, of the parameters β_0 and β_1 were obtained using the least square method by minimizing the sum of the squared residuals. According to [45,48], it is not difficult to show the following:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} x_j y_{i,j} - n\overline{xy}}{n_1 \sum_{j=1}^{n_2} x_j{}^2 - n\overline{x}^2}, \quad (2)$$

where is $n = n_1 \cdot n_2 = 39$, and that is:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}. \quad (3)$$

The value of $\overline{x}$ in expressions (2) and (3) represents the arithmetic mean of the values of the moderate scale:

$$\overline{x} = \frac{1}{n_2} \sum_{j=1}^{n_2} x_j. \quad (4)$$

It should be noticed that the values of the moderate scale are equidistant. Also, the values from the left side of zero on the moderate scale are of the opposite sign when compared to

the values from the right side of the moderate scale. That's why it's always worth that is $\bar{x} = 0$. The value of $\bar{y}$ in equations (2) and (3) is calculated from:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} y_{i,j}. \quad (5)$$

To estimate the measurement uncertainty u_0 based on the law of propagation of uncertainty (LPU) applied to the functional relationship of the input data, which was obtained from the linear regression analysis, two additional quantities are crucial: the standard error of slope $u(\hat{\beta}_1)$ and the standard error of intercept $u(\hat{\beta}_0)$. That quantities were calculated from the following equations:

$$u^2(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n_1\sum_{j=1}^{n_2} x_j{}^2 - n\bar{x}^2}\right), \quad (6)$$

and

$$u^2(\hat{\beta}_1) = \frac{\sigma^2}{n_1\sum_{j=1}^{n_2} x_j{}^2 - n\bar{x}^2}, \quad (7)$$

The estimator $\hat{\sigma}_y^2$ for σ^2 in the variable y is equal to the variance of the regression [46], so it is valid:

$$\hat{\sigma}_y^2 = \frac{1}{n-2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(y_{i,j} - y_{fj}\right)^2. \quad (8)$$

The fitted regression line for data in Table 1 has the form:

$$y_f = \hat{\beta}_0 + \hat{\beta}_1 \cdot x = -0.026154 + 1.001623 \cdot x. \quad (9)$$

The standard deviations of the slope and intercept are, respectively, $u(\hat{\beta}_1) = 0.000254$ µm and $u(\hat{\beta}_0) = 0.004752$ µm. The result for residual standard deviation in variable y is equal to $\hat{\sigma}_y = 0.029674$ µm and the coefficient of determination is $R^2 = 1$. Based on the Shapiro–Wilk test, at the significance level $\alpha = 0.05$, it was concluded that the residuals followed a normal distribution (Table S1). Residual diagnostics were also performed based on the residuals graph, Q–Q normal plot, and residual density graph (Figure S1).

2.2.2. Measurement Uncertainty

The standard uncertainty u_0 associated with the best estimate of a measurand y_0 was determined in two ways: by applying a simplified measurement model with the use of comprehensive data on measurement performance and by applying a statistical model, specifically the linear regression analysis (LRA), which was adapted to the measurement data.

A simplified measurement model that describes the measurement uncertainty of probe calibration is expressed by following equation:

$$u_c(L_P) = \sqrt{u^2(L_{ULM}) + u^2(L_{LEI}) + u^2(L_{LEII}) + u^2(L_N)}, \quad (10)$$

where $u_c(L_P)$ represents combined standard uncertainty of the probe calibration. The value denoted as $u(L_{ULM})$ is the standard uncertainty due to the influence of the accuracy of the ULM device. The marks $u(L_{LEI})$ and $u(L_{LEII})$ stand for standard uncertainty of the linear error due to the error of the probe tilt angle and for standard uncertainty of the linear error due to the error of the probe position, respectively. From experience, it is known that the error of the probe tilt angle can amount to a maximum of 5° and that the error of the probe position can amount to a maximum of 0.3 mm. The standard uncertainty due to the influence of the resolution of the measuring device is indicated with

$u(L_N)$. The components that contribute to the uncertainty of measurement results and their contributions to the combined measurement uncertainty are given in Table 2.

Table 2. Uncertainty budget for probe in the measuring range of ±30 μm [44].

Source of Uncertainty L_P	Probability Distribution	Standard Uncertainty/μm
ULM device	Normal	0.1
Linear error of the probe I	Rectangular	0.000184462
Linear error of the dial probe II	Rectangular	0.065909762
Resolution	Rectangular	0.034641016
Combined measurement uncertainty of the probe $u_c(L_P)$		0.1247

The input parameter u_0, calculated in the risk assessment procedure based on the expression for the combined measurement uncertainty [49] (p. 21), amounts to $u_0^{GUM} = 0.1247$ μm. The mark u_0^{GUM} stands for $u_c(L_P)$ in further text. The measurement uncertainty, which is calculated in this way, is the same for all points on the moderate scale **x**. All components that enter the calculation of the combined measurement uncertainty of the probe have the sensitivity coefficient $c = 1$.

Another approach to evaluating measurement uncertainty u_0 involves utilizing a statistical model and a functional relationship derived from linear regression analysis (LRA) [50,51]. The measurement uncertainty was calculated after the measurement was performed, and after the regression line was determined. This procedure encompasses the computation of the partial derivatives of the expression for the regression line and the incorporation of standard errors for the slope and intercept in the calculation of the measurement uncertainty [52,53]. Measurement uncertainty calculated using LRA is carried out for the equation:

$$Y_f = \hat{\beta}_0 + \hat{\beta}_1 X. \tag{11}$$

The uncertainty to be evaluated in the regression model is the uncertainty of the random variable Y_f evaluated for the given value $x_j, j = 1, 2, \ldots, n_2$ of the explanatory variable X. According to [49,53], these measurement uncertainties can be calculated from the equation:

$$u^2\left(y_{fj}\right) = u^2\left(\hat{\beta}_0\right) + x_j{}^2 u^2\left(\hat{\beta}_1\right) + \hat{\beta}_1^2 \sigma_x{}^2 + 2x_j u\left(\hat{\beta}_0, \hat{\beta}_1\right), j = 1, 2, \ldots, n_2, \tag{12}$$

where the standard uncertainty associated with the input quantity X is equal to

$$\sigma_x{}^2 = \frac{1}{n-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(x_j - \frac{y_{ij} - \hat{\beta}_0}{\hat{\beta}_1} \right)^2. \tag{13}$$

The numerical calculation yields a value of $\sigma_x = 0.029626$ μm for the data from Table 1. According to [52], the standard uncertainty arising from the dependence of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ equals:

$$u\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\bar{x} u^2\left(\hat{\beta}_1\right) = 0. \tag{14}$$

It is important to notice that the expression in equation (12) due to (14) only depends on the squares of the reference values of the moderate scale $x_j{}^2, j = 1, 2, \ldots, n_2$. Therefore, the graph of measurement uncertainties evaluated at the points of the moderate scale **x** is a parabola with an upward opening that is axisymmetric concerning the direction x = 0 (Figure 3).

The measurement uncertainty calculated using the expression for the combined measurement uncertainty is the same for all points of the moderate scale **x**. In contrast, the measurement uncertainties calculated by the LRA method are different for different points of the scale. The vector of measurement uncertainties, calculated by the LRA, is denoted by $\mathbf{u_0^{LRA}} = \left(u_0^1, u_0^2, \ldots, u_0^{n_2}\right)$. The point in the middle of the moderate scale has the lowest

measurement uncertainty, and the points on the edges have the highest measurement uncertainty. The values of measurement uncertainties for each point of the moderate scale **x** calculated by LRA are given in Table S2. In comparison with the value u_0^{GUM}, the values $\mathbf{u_0^{LRA}}$ are underestimated.

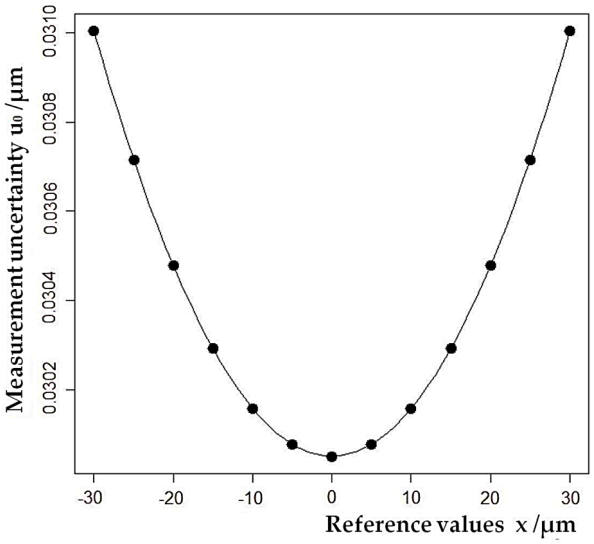


Figure 3. The measurement uncertainty of the fitted regression line according to the law of propagation of uncertainties.

When calibrating devices, there are usually no historical data from previous measurements about the slope and intercept of the regression line or their standard errors. A well-performed calibration yields a slope close to one and an intercept close to zero. For the measurer, it would be extremely difficult to determine, based on prior beliefs, what values can have a slope and an intercept, especially when it is taken into consideration that these values should be given to at least two decimal places. That kind of guessing would inevitably lead to an incorrect risk assessment. It is an even bigger problem if the data come from other non-metrology models, where the values for slope and intercept can be any. Therefore, the risk assessment problem for regression is reformulated so that the calibration data shown in Table 1 are used to determine the data y_0 and u_0 that are included in the expression for the prior distribution $g_0(\eta)$. The input parameter u_m of the likelihood function $h(\eta_m|\eta)$ is assumed to be the standard measurement uncertainty of a future inspection process. The behavior of global consumer's and producer's risk was tested for three different cases: for $u_m = u_0/2$, $u_m = u_0$ and $u_m = 2u_0$.

In that regard, the contribution of this paper consists of a proposal to indicate the data for the regression line obtained during the calibration procedure, thereby enabling the measurer to be guided by these data in future measurements. This would allow for traceability in the risk assessment.

2.2.3. Tolerance and Acceptance Interval

In the present study, four models were examined concerning the width of the tolerance interval: M1, M2, M3, and M4. The conventional tolerance interval for each point of the moderate scale was obtained based on the information provided by the manufacturer. The total span of error in both directions for the inductive contact probe in the standard range of ±30 µm amounts to 0.6 µm. Based on this information, it was assumed that in the first observed model M1, the range of tolerance interval $\Delta T = T_U - T_L$ for each point of the moderate scale **x** was $\Delta T = 0.6$ µm. Individual values of the upper and lower limits of the tolerance interval were determined for each reference value **x** of the measuring range. Through these points, a straight line could be drawn. These lines are parallel and symmetrically positioned relative to the line y = x, i.e., symmetrically relative to the line

obtained for a perfect measurement. Therefore, the upper and lower tolerance lines can be spoken about. In this sense, this model is linearized.

If information about the tolerance interval does not exist, it can alternatively be taken that the range of the tolerance interval for each point of the reference scale **x** of the moderate area is equal to $\Delta T = 4u_0^{GUM}$. Model M2, where the range of the tolerance interval is $\Delta T = 4u_0^{GUM}$, is also linearized. Model M3, where $\Delta T = 6\mathbf{u}_0^{LRA}$, was analyzed as well. In this model, the tolerance interval is placed symmetrically around the $y = x$ line, but the range of the tolerance interval at the edges of the measurement area was wider than the range of the tolerance interval in the middle of the scale. This model is non-linearized and upper and lower tolerance curves can be talked about. These tolerance interval ranges $\Delta T = 4u_0^{GUM}$ and $\Delta T = 6\mathbf{u}_0^{LRA}$ were chosen according to the 4-sigma and 6-sigma rules.

The global consumer's and producer's risk was also estimated for model M4. This model favors the minimal value of measurement uncertainty calculated using LRA. For that model, $\Delta T = 6\min(\mathbf{u}_0^{LRA})$ is valid. Model M4 was artificially linearized based on model M3. Examples of a linearized and a non-linearized model are given in Figure 4.

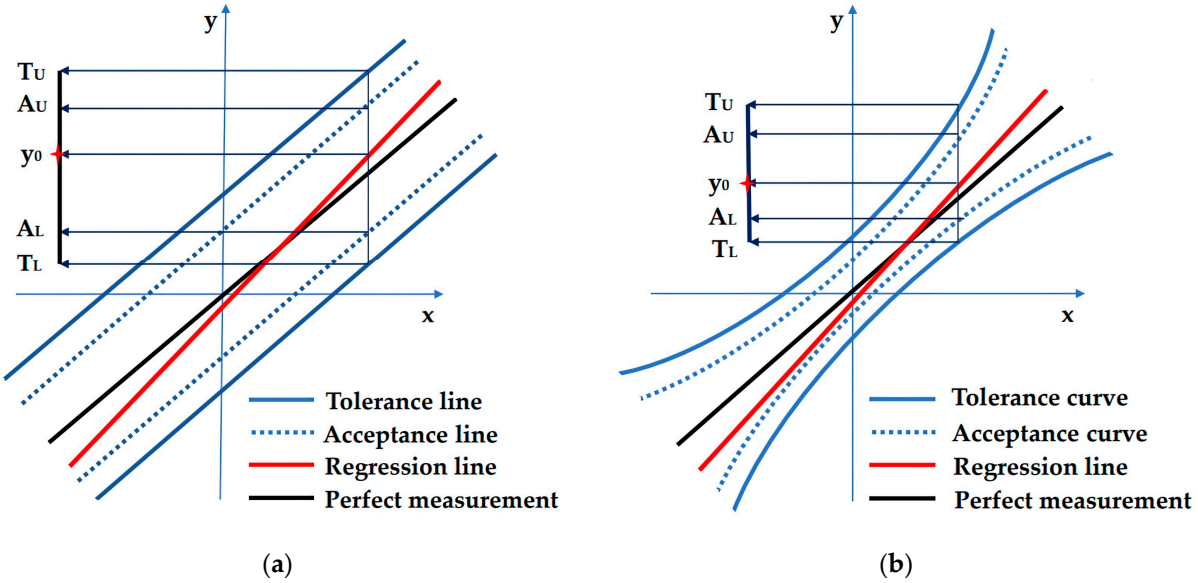


Figure 4. Spatial extension of the risk assessment model in the case of consumer risk minimization: (**a**) Linearized model; (**b**) Non-linearized model.

The primary distinction between the statistically defined simultaneous tolerance intervals and the tolerance intervals outlined in this paper lies in their construction. The tolerance intervals in this study are positioned around the $y = x$ line, which describes the perfect measurement. Simultaneous tolerance intervals are constructed relative to the fitted regression line [22]. When constructing a two-sided simultaneous tolerance interval, according to [22], for the lower and upper limits of the tolerance interval, the following are valid:

$$T_L = \hat{y}_j - \hat{\sigma}_y \cdot k_{2,j}, j = 1, 2, \ldots, n_2, \tag{15}$$

and

$$T_U = \hat{y}_j + \hat{\sigma}_y \cdot k_{2,j}, j = 1, 2, \ldots, n_2, \tag{16}$$

where k_2 is a constant which can be calculated using the formula given in [54] or using tabulated values in [55]. It follows from equations (15) and (16) that:

$$\Delta T = T_U - T_L = 2\hat{\sigma}_y \cdot k_{2,j}, j = 1, 2, \ldots, n_2. \tag{17}$$

If the simultaneous tolerance interval were to be constructed around the line $y = x$, as is the case with the other described tolerance intervals, then, due to (8), it would be valid

that $\Delta T = 0$. For this reason, simultaneous tolerance intervals cannot be applied in the risk assessment procedure for regression models defined in this paper.

If the tolerance interval is within the acceptance interval, then the maximum range of the acceptance interval $\Delta A = A_U - A_L$ is equal to $\Delta A = 1.2\Delta T$ in all models (Figure 1a). If the acceptance interval is within the tolerance interval, then the maximum range of the acceptance interval is $\Delta A = 0.8\Delta T$ (Figure 1c). For the shared risk, $\Delta A = \Delta T$ applies (Figure 1b). The range of the acceptance interval was chosen so that the global producer's risk surface would intersect with the global consumer's risk surface. For simplicity and transparency, the basic data on the analyzed models are given in Table 3.

Table 3. Overview of model properties.

Model	Uncertainty	Tolerance Interval
M1 Linearized model	$u_0 = u_0^{GUM}$	$\Delta T = 0.6\ \mu m$
M2 Linearized model	$u_0 = u_0^{GUM}$	$\Delta T = 4u_0^{GUM}$
M3 Non-linearized model	$u_0 = \mathbf{u_0}^{LRA}$	$\Delta T = 6\mathbf{u_0}^{LRA}$
M4 Linearized model	$u_0 = \mathbf{u_0}^{LRA}$	$\Delta T = 6\min(\mathbf{u_0}^{LRA})$

Finally, it should be emphasized that the risk assessment was carried out for the regression line even though the data from Table 1 were collected during the calibration procedure. During the calibration procedure, the measurement uncertainty determined according to the LRA, instead of for $y_{fj}, j = 1, 2, \ldots, n_2$ from equation (11), would be determined for the values of the moderate scale **x**. The equation for variable x is obtained from the expression (11) for the regression line [53] (p. 872). Furthermore, to calculate the risk for the calibration procedure, it is necessary to invert the tolerance and acceptance intervals constructed for the regression line to obtain the tolerance and acceptance intervals for the explanatory variable [56] (p. 29).

2.3. Risk Calculation

This paper observes the behavior of the global producers' risk R_P and global consumers' risk R_C at points defined along the guard band axis and along the moderate scale. As a result of the analysis, according to the procedure described below, risk surfaces were obtained.

It is assumed that the maximal length of the guard band at each point of the moderate scale **x** is equal to

$$w_{max_j} = 0.2\left(T_{U_j} - T_{L_j}\right), j = 1, 2, \ldots, n_2, \tag{18}$$

where the value of the upper limit of the tolerance interval at the points of the moderate scale **x** is given by the following equation:

$$T_{U_j} = y_j + \frac{\Delta T}{2}, j = 1, 2, \ldots, n_2, \tag{19}$$

and the value of the lower limit of the tolerance interval at points of the moderate scale **x** is given by the following equation:

$$T_{L_j} = y_j - \frac{\Delta T}{2}, j = 1, 2, \ldots, n_2. \tag{20}$$

In equations (19) and (20), the values for ΔT from Table 3 are included. In addition, it is valid that the points $y_j = y(x_j) = x_j$ are the points of the line $y = x$, which corresponds to the perfect measurement. In other words, from equation (18), it follows that the acceptance interval occupies 80% of the tolerance interval. The 10% of the length of the guard band

occupies the area between the acceptance line (curve) A_L and the tolerance line (curve) T_L, below the line $y = x$, and 10% of the length of the guard band occupies the area between the acceptance line (curve) A_U and the tolerance line (curve) T_U, above the line $y = x$ (Figure 4).

The maximal length of the guard band $w_{max_j}, j = 1, 2, \ldots, n_2$ in linearized models is the same for all points of the moderate scale **x**. For the non-linearized model M3, the maximal length of the guard band w_{max_j} has the smallest value at the middle point of the moderate scale, for $x_7 = 0$. The w_{max_j} in non-linearized models has the highest value at the edges of the moderate scale.

The lower and upper limits of the acceptance interval are defined by introducing a multiplicative factor $r \in [-1, 1]$. An equidistant subdivision of the interval $[-1, 1]$ with a subdivision rate of 0.1 results in 21 subdivision nodes [18]. These subdivision nodes have the form $r_k = -1 + 0.1 \cdot (k - 1), k = 1, 2, \ldots, n_3$, where $n_3 = 21$. For nodes of the subdivision r_k, the lower and upper limits of the acceptance interval at the points of the reference scale **x** are calculated from the following formulas:

$$A_{L_{j,k}} = T_{L_j} + \frac{r_k \cdot w_{max_j}}{2}, j = 1, 2, \ldots, n_2, k = 1, 2, \ldots, n_3, \tag{21}$$

and

$$A_{U_{j,k}} = T_{U_j} - \frac{r_k \cdot w_{max_j}}{2}, j = 1, 2, \ldots, n_2, k = 1, 2, \ldots, n_3. \tag{22}$$

For $k = 1$, it is valid that $r_k = -1$, and in that case, $\Delta A_j = 1.2\left(T_{U_j} - T_{L_j}\right), j = 1, 2, \ldots, n_2$, i.e., the tolerance interval is within the acceptance interval (Figure 1a). For $k = 21$, it is valid that $r_k = 1$. In this case, $\Delta A_j = 0.8\left(T_{U_j} - T_{L_j}\right), j = 1, 2, \ldots, n_2$, and the acceptance interval is within the tolerance interval (Figure 1c). For $k = 11$, it is valid that $r_k = 0$ and that $\Delta A_j = \left(T_{U_j} - T_{L_j}\right) = \Delta T, j = 1, 2, \ldots, n_2$. In this case, it is a shared risk model (Figure 1b).

It should be noted that for all models, the tolerance lines (curves) calculated from equations (19) and (20) are fixed. The acceptance lines (curves) change along the guard band axis, depending on the length w_k of the guard band:

$$w_k = \frac{r_k \cdot w_{max}}{2}, k = 1, 2, \ldots, n_3. \tag{23}$$

According to [13], the risk of rejection of a product conforming to the specifications, i.e., the global producers' risk R_P, is calculated from the equation:

$$R_P = \int_{-\infty}^{A_L} \int_{T_L}^{T_U} g_0(\eta) h(\eta_m|\eta) d\eta_m d\eta + \int_{A_U}^{\infty} \int_{T_L}^{T_U} g_0(\eta) h(\eta_m|\eta) d\eta_m d\eta. \tag{24}$$

The risk of accepting a non-compliant product, i.e., the global consumer's risk R_C, according to [13], is calculated using the following equation:

$$R_C = \int_{-\infty}^{T_L} \int_{A_L}^{A_U} g_0(\eta) h(\eta_m|\eta) d\eta_m d\eta + \int_{T_U}^{\infty} \int_{A_L}^{A_U} g_0(\eta) h(\eta_m|\eta) d\eta_m d\eta. \tag{25}$$

The random variables Y and Y_m are normally distributed. Therefore, the prior $g_0(\eta)$ in equations (24) and (25) has the form:

$$g_0(\eta) = \frac{1}{u_0\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\eta - y_0}{u_0}\right)^2\right], \tag{26}$$

and the likelihood function has the form:

$$h(\eta_m|\eta) = \frac{1}{u_m\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\eta_m - \eta}{u_m}\right)^2\right]. \tag{27}$$

In the simplification procedure, which is elaborately outlined in detail in [18], the double integrals present in formulas (24) and (25) can be reduced to single integrals. The global producer risk for models M1 and M2 can be calculated using the expression:

$$R_{P_{j,k}} = \int_{\frac{T_{L_j} - y_f(x_j)}{u_0{}^{GUM}}}^{\frac{T_{U_j} - y_f(x_j)}{u_0{}^{GUM}}} \varphi_0(z)\left(1 - F(z)_{j,k}\right)dz, j = 1, 2, \ldots, n_2, k = 1, 2, \ldots, n_3, \quad (28)$$

where the label φ_0 denotes density function for the unit normal distribution, which can be calculated from the following equation:

$$\varphi_0(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right), \quad (29)$$

The quantity $F(z)_{j,k}, j = 1, 2, \ldots, n_3$, $k = 1, 2, \ldots, n_3$ is expressed through the cumulative distribution function (CDF) for the unit normal distribution [13]. For models M1 and M2, $F(z)_{j,k}$ has the form:

$$F(z)_{j,k} = \phi\left(\frac{A_{U_{j,k}} - y_f(x_j) - z{\cdot}u_0{}^{GUM}}{u_m}\right) - \phi\left(\frac{A_{L_{j,k}} - y_f(x_j) - z{\cdot}u_0{}^{GUM}}{u_m}\right), j = 1, 2, \ldots, n_2, k = 1, 2, \ldots, n_3 \quad (30)$$

Global consumers' risk is calculated, according to [13], from

$$R_{C_{j,k}} = \int_{-\infty}^{\frac{T_{L_j} - y_f(x_j)}{u_0{}^{GUM}}} \varphi_0(z)F(z)_{j,k}dz + \int_{\frac{T_{U_j} - y_f(x_j)}{u_0{}^{GUM}}}^{\infty} \varphi_0(z)F(z)_{j,k}dz, j = 1, 2, \ldots, n_2, k = 1, 2, \ldots, n_3. \quad (31)$$

For the standard measurement uncertainty of future measurements u_m, in equations (27) and (30), the values $u_m = u_0/2$, $u_m = u_0$ and $u_m = 2u_0$ were taken. The formulas for global producer's risk and global consumer's risk for the non-linearized model M3 and the linearized model M4 were obtained by inserting the values $u_{0_j}{}^{LRA}, j = 1, 2, \ldots, n_2$ instead of $u_0{}^{GUM}$ into equations (28), (30) and (31).

It is clear from equations (18)–(31) that the global risks of producers and consumers are calculated for each value of the moderate scale, namely, for all reference values $x_1, x_2, \ldots, x_{n_2}$ and for all values of the multiplicative factor $r_k, k = 1, 2, \ldots, n_3$ in the nodes of the subdivision of the guard band. In this way, risk surfaces are obtained. The total number of grid points where the values of global producers' and consumers' risk were evaluated is $n_2{\cdot}n_3 = 273$. The risk evaluation was performed by numerically solving the integrals from equations (28) and (31). All calculations and 2D graphs were performed by using the R software package, version 4.2.0, while 3D graphs were created using the Octave software package, version 8.4.0 [57–59]. This can be a challenge because the measurer is required to be familiar with mathematics and programming. That is why risk calculation is often avoided.

An important quantity when assessing the global risk of consumers, R_c, and the global risk of producers, R_P, which serves as a measure of model quality, is conformance probability, p_C. This quantity indicates the probability that the measured value of the item of interest is within the tolerance interval. According to [13] (p. 27) and equation (26), conformance probability for models M1 and M2 can be calculated from:

$$p_{C_j} = \frac{1}{u_0{}^{GUM}\sqrt{2\pi}}\int_{T_{L_j}}^{T_{U_j}} \exp\left[-\frac{1}{2}\left(\frac{\eta - y_f(x_j)}{u_0{}^{GUM}}\right)^2\right]d\eta, j = 1, 2, \ldots, n_2. \quad (32)$$

For the models M3 and M4, conformance probability is calculated by including the value $u_{0_j}{}^{LRA}, j = 1, 2, \ldots, n_2$ in equation (32), instead of the value $u_0{}^{GUM}$.

3. Results and Discussion

3.1. Graphical Risk Analysis

Graphical risk analysis was carried out by monitoring the behavior of the global risk of the producer R_P and the global risk of the consumer R_C along the moderate scale and along the guard band axis and monitoring the behavior of the risk concerning the assumed value of measurement uncertainty u_m of a future inspection process.

3.1.1. Behaviors of a Global Consumer and Producer Risk along the Moderate Scale

The behavior of the global producer's risk R_P and global consumer's risk R_C along the moderate scale depends on the values of the slope $\hat{\beta}_1$ and intercept $\hat{\beta}_0$ of the regression line. The risk curves for R_P and R_C along the moderate scale are parabolas with upward openings. The minimum of these parabolas is located at the intersection of the regression line $y_f = \hat{\beta}_0 + \hat{\beta}_1 x$ and the line y = x. The position of the minimum at the moderate scale is indicated by x_{min} and is represented by the following expression:

$$x_{min} = \frac{\hat{\beta}_0}{1 - \hat{\beta}_1}. \tag{33}$$

It is easy to show how the minimal values of the global risk of producers and consumers along the moderate scale in all analyzed models are achieved for $x_{min} = 16.12$ µm (Figure 5).

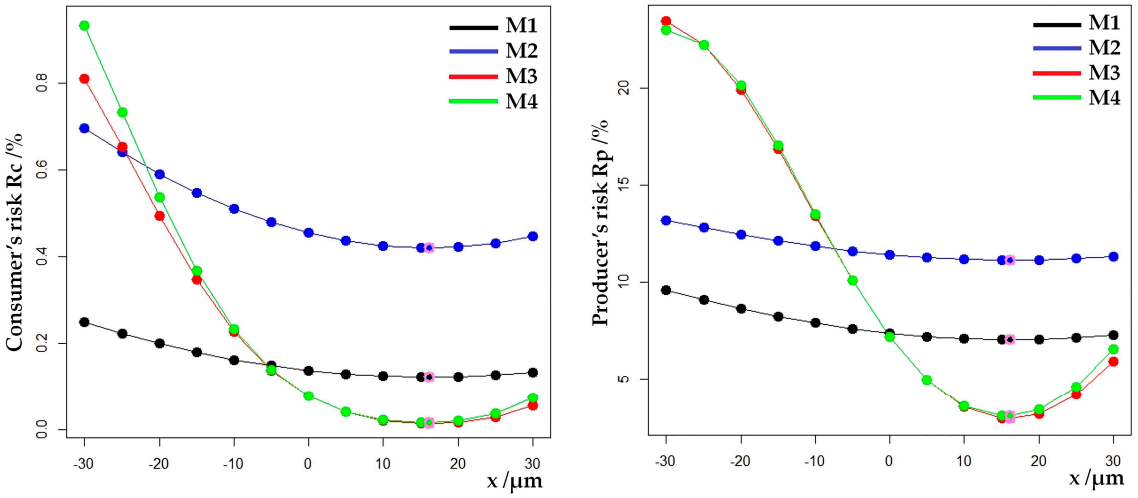


Figure 5. Comparison of behavior of consumer's risk and producer's risk along the moderate scale for all four models, $r = 1$, $u_m = u_0/2$: (**a**) Consumer's risk; (**b**) Producer's risk. A violet circle indicates the point of minimum for the risks R_C and R_P.

Parabolas are translated to the right when $x_{min} > 0$, i.e., if $\hat{\beta}_0 > 0$ and $\hat{\beta}_1 < 1$, or if, as is the case here, $\hat{\beta}_0 < 0$ and $\hat{\beta}_1 > 1$. That is the reason why the risk parabolas, i.e., the iso-risk curves along the moderate scale depicted in Figure 5, are asymmetric. In all models, due to translation, the values of global risk for producers and consumers on the left side of the moderate scale are higher compared to the risk values on the right side of the moderate scale. This disparity is readily apparent in Figure 4, as evidenced by the proportion of the length of the regression line located on the right side of the intersection of the regression line and the $y = x$ line, in contrast to the proportion located on the left side of the intersection of the regression line and the $y = x$ line. Therefore, the calculated global

risks of producers and consumers can be seen as risks of deviation of the regression line from the $y = x$ line.

In the case of a regression line where $\hat{\beta}_0 > 0$ and $\hat{\beta}_1 > 1$, or $\hat{\beta}_0 < 0$ and $\hat{\beta}_1 < 1$, the parabolas would be translated to the left. The minimal values for the parabolas of the global risk of producers and consumers would then be located at the point of the moderate scale $x_{min} < 0$, and would be calculated from equation (33).

The line marked with y_s and given by the following equation:

$$y_s = -\hat{\beta}_0 + (2 - \hat{\beta}_1) \cdot x, \tag{34}$$

is axisymmetric to the regression line. These lines are axisymmetric concerning $y = x$. For the $\hat{\beta}_1 > 1$ worth $2 - \hat{\beta}_1 < 1$, and conversely, for the $\hat{\beta}_1 < 1$ worth $2 - \hat{\beta}_1 > 1$. The global risks of producers and consumers calculated for the line y_s are equal to the risks calculated for the regression line from equation (9) up to the order of magnitude ranging from 10^{-4} to 10^{-15}, depending on the observed model and assuming that the line y_s is obtained under identical conditions as the regression line, with identical values for u_0 and u_m and identical values for tolerance intervals and acceptance intervals. The same applies to conformance probability.

Due to the narrower range of the tolerance interval ΔT, the M2 model exhibits higher values regarding the global risk of producers and consumers across the entire moderate scale compared to the M1 model. The width of the tolerance interval of the M3 model is greater than the width of the tolerance interval of the M4 model, except in the middle of the scale, at the point $x_7 = 0$. At this point on the moderate scale, the widths of the tolerance intervals for both models are equal, amounting to 6 min $\left(\mathbf{u}_0^{LRA}\right)$ (Table 3). Consequently, the global risks of producers and consumers in the M3 model are lower than those in the M4 model (Figure 5a,b). In these two models, the risk values coincide at the point of minimum, x_{min}.

The iso-risk curves of models M1 and M2, along the moderate scale, exhibit stable behavior with a narrower range of value changes. These models behave differently compared to the M3 and M4 models. In models M3 and M4, the range of variations in producer and consumer risk values is significantly greater. In the case of consumer risk, the iso-risk curves for models M3 and M4 intersect those for models M1 and M2 (Figure 5a). The same applies to iso-risk curves of producers (Figure 5b). All models are shown on the same scale for comparison. That is why the iso-risk curves for models M1 and M2 do not have such prominent parabola shapes as models M3 and M4.

The maximal value for conformance probability is reached precisely at the point x_{min}, where the global risk of the producer R_P and the global risk of the consumer R_C have a minimal value. Higher risk values result in lower values for the conformance probability and vice versa. This can be observed when comparing models M1 and M2, i.e., M3 and M4, as well as all models (Figure 6). Models M3 and M4 have the same value for conformance probability at the point of moderate scale, x_{min}. Other points of the moderate scale show higher values for conformance probabilities for model M3 than for model M4. The lowest values for conformance probability are found on the left side of the moderate scale for all models. The values for p_C drop to 72% for the M3 model, and even to 69% for the M4 model. Hence, it is possible to conclude that models M3 and M4 attach greater importance to risks at the edges of the scale.

On the graphs of models M3 and M4, which show the behavior of the global risk of the producers, anomalies on the left edge of the scale can be observed (Figure 5b). Anomalies mean that the iso-risk curves deviate from the parabola graph. At the point of the moderate scale $x_1 = -30$ µm, the global risk of the producer for the M3 model is greater than the global risk of the producer for the M4 model. These anomalies occur when the tolerance interval is too narrow. If the range of the tolerance interval is too narrow, the regression line crosses the acceptance line, or both the acceptance and tolerance lines (or acceptance and tolerance curves). In that case, the risk graphs are no longer parabolas (Figure 7).

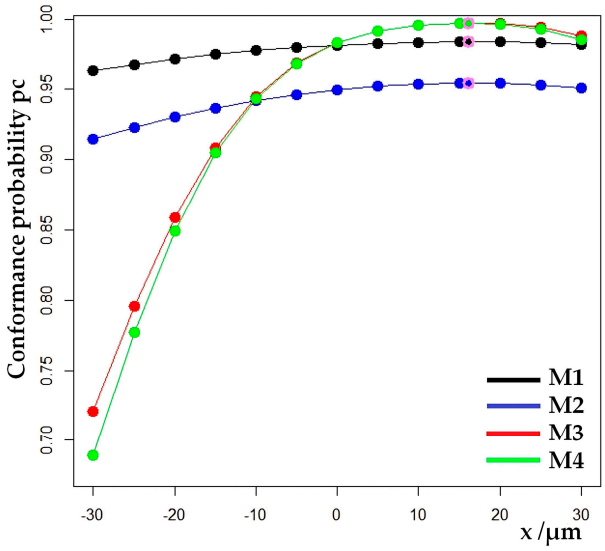


Figure 6. Comparison of behavior of conformance probability along the moderate scale for all four models, $r = 1$, $u_m = u_0/2$. A violet circle indicates the point of maximum for conformance probability p_C.

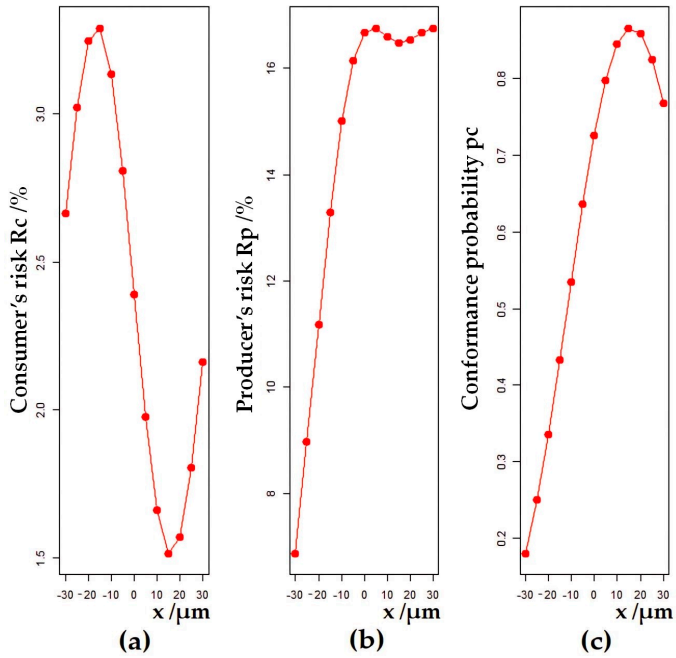


Figure 7. An example of anomalies in the behavior of global consumers' and producers' risks along the moderate scale for the case of ${u_0}^{LRA}$, $\Delta T = 3\mathbf{u_0}^{LRA}$, $r = -1$, $u_m = u_0/2$: (**a**) Consumer's risk; (**b**) Producer's risk; (**c**) Conformance probability.

Figure 7 illustrates the presence of anomalies, i.e., the deviations of the iso-risk curve from the parabola graph. Such deviations are obtained when, for example, with the M3 model, a too-narrow tolerance interval range of $\Delta T = 3\mathbf{u_0}^{LRA}$ is set. In that scenario, the regression line intersects with the lower acceptance curve A_L and the lower tolerance curve T_L. Anomalies may manifest themselves for both the global consumer's risk and global producer's risk (Figure 7a,b). The graph for conformance probability clearly shows that the model behaves badly on the left side of the moderate scale, where the conformance probability falls below 20% (Figure 7c). The primary drawback of the described method for risk assessment in regression is the excessively narrow tolerance interval. It is necessary to expand its span to resolve the issue. To stay within the 6-sigma range used in statistics,

the tolerance interval width was set to $\Delta T = 6\mathbf{u_0}^{LRA}$ in the M3 model. The anomaly was not completely resolved by this. The regression line for $r \approx 0.9770$ intersected the lower acceptance curve A_L. The iso-risk curves for the global producer's risk R_P diverged from the parabola graph on the left edge of the moderate scale (Figure 5b). For this reason, the range of tolerance intervals should be further expanded. The deviation from the parabola graph was even more noticeable with the M4 model. With this model, the regression line intersected the lower line of the acceptance interval A_L already by $r \approx 0.8497$. In both models, M3 and M4, anomalies occur for the values of the multiplicative factor r which is related to the model of minimization of global consumer's risk (Figure 1c). The specified values of the multiplicative factor r were calculated numerically. Also, although the regression lines in models M3 and M4 intersect the lower acceptance curve (line) A_L, they fail to intersect the lower tolerance curve (line) T_L. For the M2 model, it is sufficient to set $\Delta T = 4u_0{}^{GUM}$ to prevent the regression line from intersecting either the tolerance interval or acceptance interval.

3.1.2. Behaviors of Global Consumer and Producer Risk along the Guard Band Axis

Depending on the multiplicative factor r_k, where $k = 1, 2, \ldots, n_3$, according to equation (24), the guard band w_k can have a positive or negative value or be equal to zero. If $w_k < 0$, it is a model of the minimization of global producer's risk (Figure 1a). If $w_k = 0$, it is a shared risk model (Figure 1b). For $w_k > 0$, it is a model for minimizing the global consumer risk (Figure 1c). It is natural to observe the behavior of the global risk of producers and consumers along the guard band axis. In general, if it is $w_{k_1} < w_{k_2}$ then for $k_1 \neq k_2, k_1 < k_2$ and $k_1, k_2 \epsilon \{1, 2, \ldots, n_3\}$, it holds that $R_{C_{k_1}} > R_{C_{k_2}}$ and $R_{P_{k_1}} < R_{P_{k_2}}$. Simplified, the consumer's risk decreases along the guard band axis as depicted in Figure 8a, whereas the producer's risk increases along the guard band axis as depicted in Figure 8b.

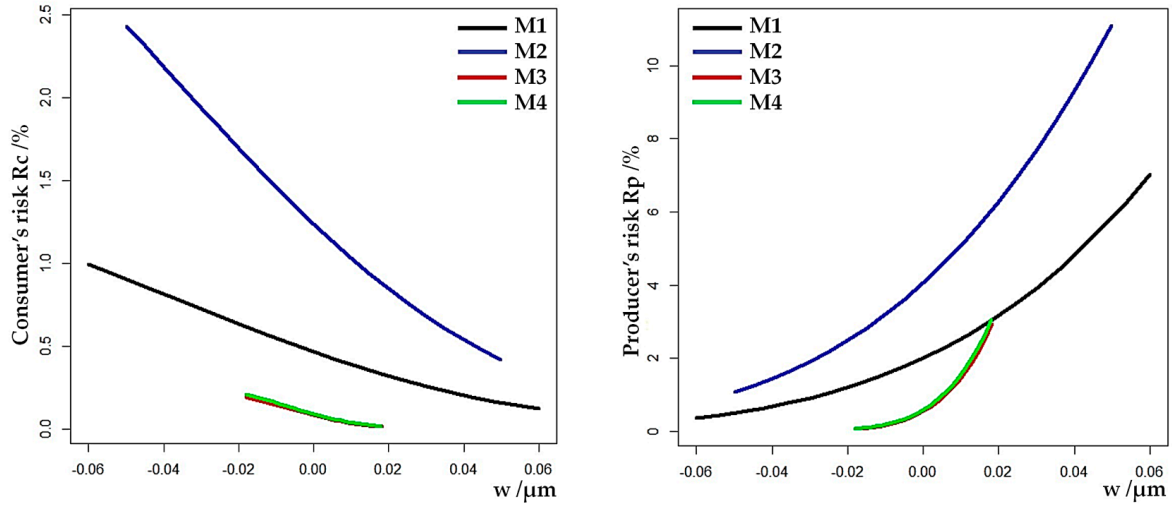


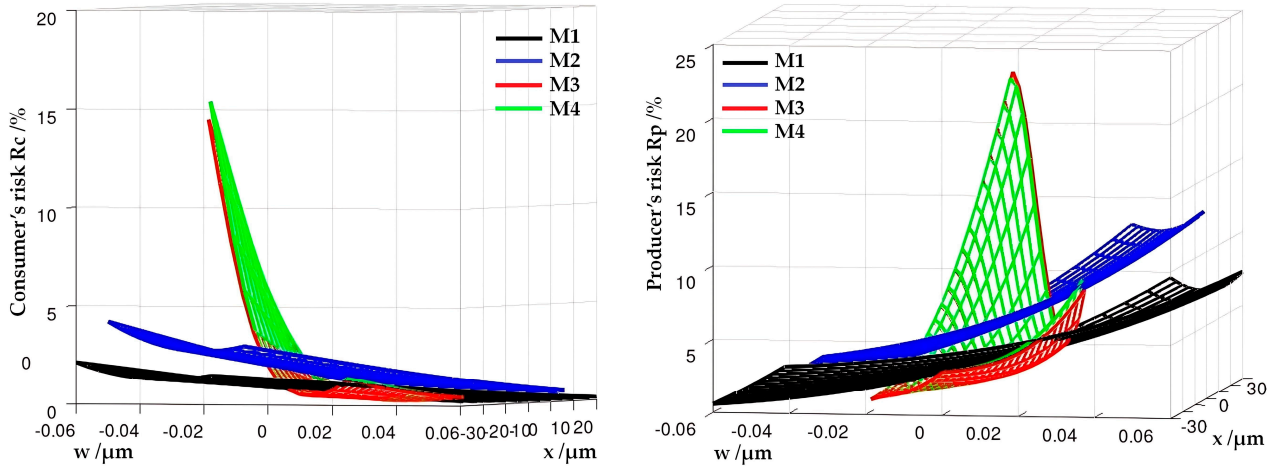
Figure 8. Curves of a minimum of the global risk of consumers and producers evaluated along the guard band axis at points $(x_{min}, w_k), k = 1, 2, \ldots, n_3$, $r = 1$, $u_m = u_0/2$: (**a**) Consumer's risk; (**b**) Producer's risk.

Figure 8 shows the curves of the minimum of the global producer's and consumer's risk calculated numerically along the guard band axis at the points $(x_{min}, w_k), k = 1, 2, \ldots, n_3$. These are the curves where the risks for each model are the least. Due to the described construction of acceptance and tolerance intervals, risks were assessed for intervals of varying lengths along the guard band axis. The values of measurement uncertainties calculated using LRA, $\mathbf{u_0}^{LRA}$, are considerably smaller in comparison with u_0^{GUM}. It is, therefore, pointless to compare all models for the same tolerance interval, $\Delta T = 0.6$ µm,

given by the manufacturer. For the measurement uncertainty $\mathbf{u}_0^{LRA}$, the tolerance interval selected in this manner would be too wide, resulting in negligible risk values R_C and R_P for models M3 and M4, while the conformance probability would be equal to one. The M4 model's curve of minimum for the global producer's risk intersects the curve of minimum of the M1 model (Figure 8b).

The conformance probability for the curve of minimum in each model is determined by the straight line along the guard band axis. These values remain the same for all points $(x_{min}, w_k), k = 1, 2, \ldots, n_3$. For the M1, M2, M3, and M4 models, respectively, they are $p_C \approx 0.9838$, $p_C \approx 0.9545$, $p_C \approx 0.9973$, and $p_C \approx 0.9971$.

The described behavior of the global producer's risk R_P and the global consumer's risk R_C along the moderate scale and the guard band axis can be easily noticed by observing the risk surfaces shown in Figure 9. From the picture, it is evident that there are areas where the global risk of the producer and the global risk of the consumer of models M3 and M4 are lower than the risk of models M1 and M2.



(**a**) (**b**)

Figure 9. Risk surfaces: (**a**) Consumer's risk; (**b**) Producer's risk.

The conformance probability curves along the moderate scale are as shown in Figure 6, but the conformance probability surfaces were evaluated for intervals of different ranges along the guard band axis (Figure 10).

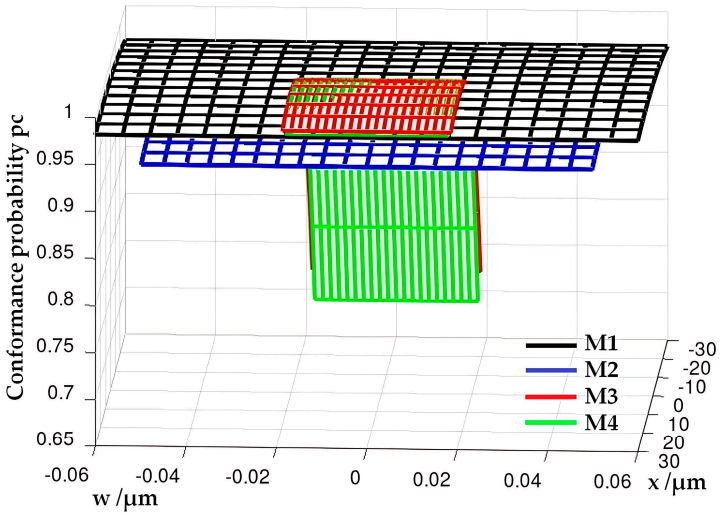


Figure 10. Conformance probability surfaces.

3.1.3. Behavior of Global Consumer's and Producer's Risk with the Changes in Measurement Uncertainty u_m of a Future Inspection Process

The behavior of the global producer's risk R_P and the global consumer's risk R_c across all models were tested for three different assumed values of measurement uncertainty u_m of a futured inspection process: for $u_m = u_0/2$, $u_m = u_0$, and $u_m = 2u_0$.

With the increase in measurement uncertainty u_m, the global consumer's risk R_C also increases (Figure 11).

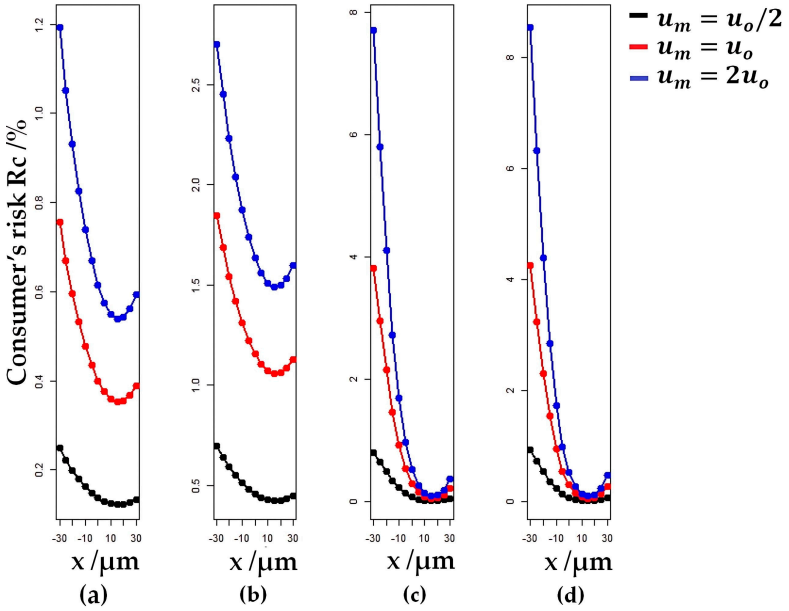


Figure 11. Behaviors of global consumer's risk along the moderate scale with the changes in measurement uncertainty u_m, r = 1: (**a**) Model M1; (**b**) Model M2; (**c**) Model M3; (**d**) Model M4.

Additionally, with the increase in the measurement uncertainty value u_m, the producer's risk R_P also rises (Figure 12).

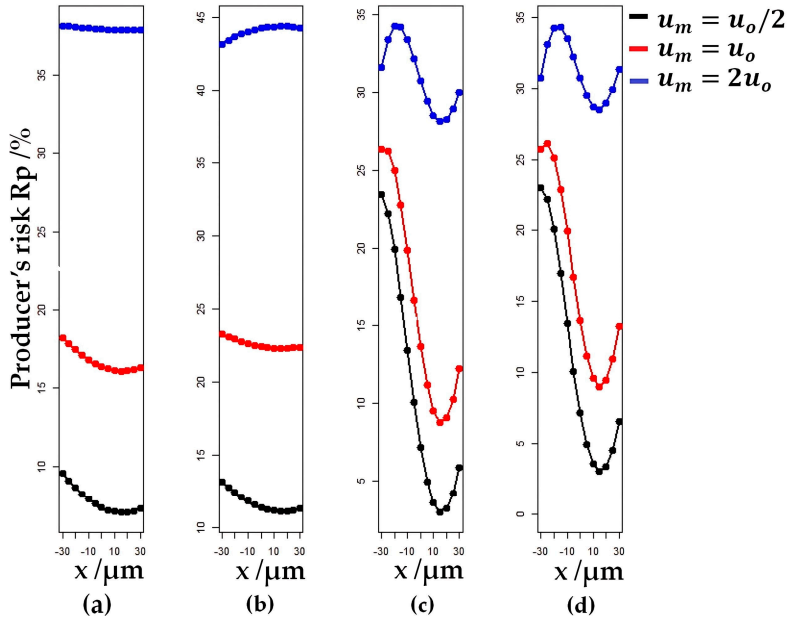


Figure 12. Behaviors of global producer's risk along the moderate scale with the changes in measurement uncertainty u_m, r = 1: (**a**) Model M1; (**b**) Model M2; (**c**) Model M3; (**d**) Model M4.

It is evident that, in addition to the range of the tolerance interval to the occurrence of anomalies in the behavior of the global producer's risk R_P, measurement uncertainty u_m also has an impact. Deviation from the parabolic curve becomes more pronounced (Figure 12c,d). Graphs can completely alter their behavior, as is the case for model M2. When $u_m = 2u_0$, the graph of the global producer's risk R_P is a parabola with a downward opening (Figure 12b). For $u_m = 2u_0$, the regression line is not outside the tolerance interval or the acceptance interval in any of the mentioned cases. To avoid anomalies in the behavior of the risk curves for R_P, it is necessary to expand the ranges of tolerance intervals. For model M2, this range should be set to at least $\Delta T = 4.4{u_0}^{GUM}$; for model M3, to $\Delta T = 7.4\mathbf{u_0}^{LRA}$; and for model M4, to $\Delta T = 7.7\mathbf{u_0}^{LRA}$. This prevents anomalies in the graphs for all values of $r_k, k = 1, 2, \ldots, n_3$.

According to equation (32), the expression for conformance probability p_C is independent of the measurement uncertainty u_m of a future inspection process. Therefore, the graphs for conformance probability for all tested values of measurement uncertainty u_m are as shown in Figure 6, and for the 3D case, as shown in Figure 10.

3.2. Comparison of Models by Root Mean Squared Error

There are many evaluation metrics for regression [60,61]. The most common are the mean squared error (MSE) and the root of the mean squared error (RMSE). RMSE is a more suitable statistical indicator compared to MSE because it is measured in the same units as the targeted variable, which allows for easier interpretation and comparison of models [62]. RMSE values range from zero to infinity. Lower RMSE values, closer to zero, indicate better model performance [63]. This metric is sensitive to large deviations of the measured values from the reference values [64].

The comparison of risk models using the RMSE metric was conducted both graphically and quantitatively. The targeted variables in the model comparison were global producer's and consumer's risk and conformance probability. Comparisons were conducted for each point on the moderate scale $\mathbf{x}$ and for all measurements $\mathbf{y_i}, i = 1, 2, 3$. If $R_{C_{j,k|y_i}}, i = 1, 2, 3$ denotes the global consumer's risk, calculated for each realization of the sample $\mathbf{y_i}, i = 1, 2, 3$, then the RMSE for the global consumer's risk, for a chosen fixed value $j \in \{1, 2, \ldots, n_2\}$, can be calculated according to the following equation:

$$\mathrm{RMSE_R}_{C_{y_i}}(j) = \sqrt{\frac{1}{n_3}\left[\sum_{k=1}^{n_3}\left(R_{C_k}(j) - R_{C_{k|y_i}}(j)\right)^2\right]}, k = 1, 2, \ldots, n_3. \tag{35}$$

Analogously, by introducing the notations $R_{P_{j,k|y_i}}$ and $p_{C|y_i}, i = 1, 2, 3$, the expressions for the RMSE of the individual measurements $\mathbf{y_i}, i = 1, 2, 3$ can be obtained for the global producer's risk and conformance probability, respectively. These metrics quantify the deviation of the calculated risk for each measurement point from the fitted risk surfaces shown in Figure 9, or from the fitted surfaces for conformance probability shown in Figure 10.

The RMSE metric exhibits higher values for those points on the moderate scale for which the deviation of measured values is greater relative to the scale values (Figure 13). The calculated RMSE values indicate that models M3 and M4 better detect deviations of the risks, calculated for the points of the moderate scale, from the risk surfaces. This is particularly true for the sample realization $\mathbf{y_1}$ on the negative part of the scale (Table 1). All the points on the moderate scale, for models M1 and M2, for each measurement $\mathbf{y_i}, i = 1, 2, 3$, and for each value of $k = 1, 2, \ldots, n_3$, fall within the tolerance interval and acceptance interval. The same holds for all models, for measurements $\mathbf{y_i}, i = 2, 3$ and for each value of $k = 1, 2, \ldots, n_3$. For model M3, the value $y_{1,2} = -25.12$ µm is already outside of the tolerance interval and acceptance interval for $k = 1$. As consequences, there is a higher producer's and consumer's risk for the mentioned measurement, and ultimately, higher RMSE values for that measurement. For $k = 1$, all other values of the negative portion of the moderate scale $\mathbf{x}$ are outside the tolerance interval, that is, between the tolerance

interval and the acceptance interval. From k = 8 to k = 21, all six measurements on the negative part of the moderate scale from sample y_1 are outside the tolerance interval and outside the acceptance interval. In the model M4, all points in the negative portion of the moderate scale are located outside the tolerance and acceptance intervals for k = 6 to k = 21.

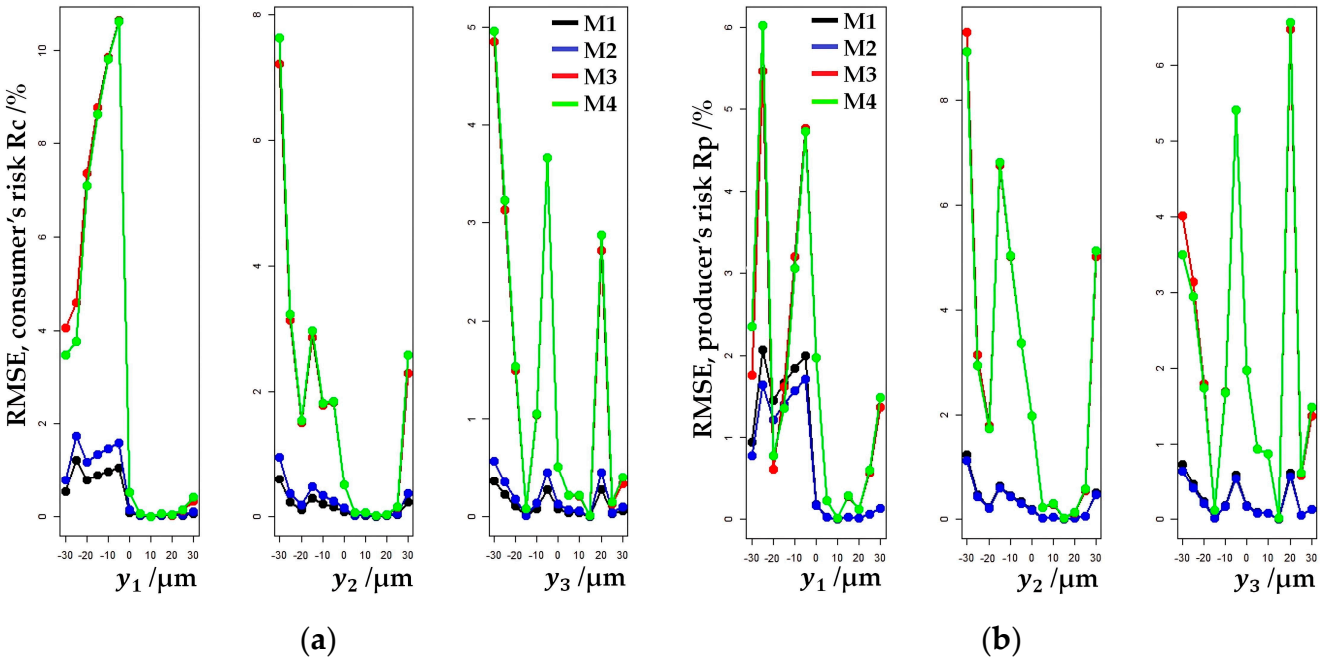


Figure 13. The comparison of the RMSE values of each individual model, calculated in the values of the moderate scale **x**, for each sample realization $\mathbf{y_i}$, i = 1, 2, 3: (**a**) Consumer's risk; (**b**) Producer's risk.

Analogously, the RMSE values for conformance probability are higher in points where the measured values significantly diverge from the values of the moderate scale, in contrast to measurements with minor deviations from the moderate scale (Figure S2).

The total value of the RMSE for each model M_l, l = 1, 2, 3, 4 for the consumer's risk, calculated for each measurement $\mathbf{y_i}$, i = 1, 2, 3, is denoted as $R_C_M_l$, l = 1, 2, 3, 4 and obtained from the equation:

$$R_C_M_l = \sqrt{\frac{1}{n_2 n_3}\left[\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\left(R_{C_k}(j) - R_{C_{k|y_i}}(j)\right)^2\right]}, l = 1, 2, 3, 4. \quad (36)$$

Expressions for the total value of RMSE for the global producer's risk and conformance probability are denoted as $R_P_M_l$ and $p_C_M_l$, l = 1, 2, 3, 4, respectively. The values of these quantities are calculated analogously to the equation (36).

Each measurement $\mathbf{y_i}$, i = 1, 2, 3 contributes to the determination of the parameters of the regression line and the determination of the risk surfaces and surfaces for the conformance probability [46,47] (pp. 76–78). Therefore, the total RMSE value for global consumer's risk was additionally calculated for each model. This value is labeled as $total_R_c$, and was calculated according to the following equation:

$$total_R_C = \sqrt{\frac{1}{n_1 n_2 n_3}\left[\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\left(R_{C_k}(j) - R_{C_{k|y_i}}(j)\right)^2\right]}. \quad (37)$$

The total value of RMSE for the global producer's risk, $total_R_P$, and the total value of RMSE for conformance probability, $total_p_c$, are obtained in the same manner as in

equation (37). The results of the quantitative analysis for all models, considering all defined values for RMSE given in equations (35)–(37), are presented in Table 4.

Table 4. RMSE values for global consumer's and producer's risk and conformance probability according to models and sample realization, and in total.

RMSE, Global Consumer's Risk				
Measurement y_i/µm	**R_C_M_1/%**	**R_C_M_2/%**	**R_C_M_3/%**	**R_C_M_4/%**
$\mathbf{y_1}$	0.63	0.94	5.40	5.27
$\mathbf{y_2}$	0.22	0.35	2.55	2.68
$\mathbf{y_3}$	0.17	0.27	2.11	2.16
*total*_R_c	0.40	0.60	3.66	3.63
RMSE, Global Producer's Risk				
Measurement y_i/µm	**R_P_M_1/%**	**R_P_M_2/%**	**R_P_M_3/%**	**R_P_M_4/%**
$\mathbf{y_1}$	1.16	0.97	2.40	2.52
$\mathbf{y_2}$	0.46	0.42	4.03	3.97
$\mathbf{y_3}$	0.35	0.32	2.92	2.87
*total*_R_P	0.75	0.64	3.19	3.18
RMSE, Conformance Probability				
Measurement y_i/µm	**p_C_M_1**	**p_C_M_2**	**p_C_M_3**	**p_C_M_4**
$\mathbf{y_1}$	0.022	0.039	0.356	0.358
$\mathbf{y_2}$	0.007	0.014	0.092	0.100
$\mathbf{y_3}$	0.005	0.010	0.078	0.083
*total*_p_c	0.013	0.025	0.217	0.220

Quantitative analysis revealed that models M3 and M4 are better at detecting deviations of measured values from values on a moderate scale. Consequently, all RMSE values for these models are higher compared to models M1 and M2. For all individual measurements $\mathbf{y_i}$, i = 1, 2, 3, it holds that R_C_$M_1 < R_C$_M_2, which is understandable given that model M1 has a broader tolerance interval compared to model M2. Furthermore, for RMSE values associated with the producer's risk, it holds that R_P_$M_1 > R_P$_M_2 for all measurements $\mathbf{y_i}$, i = 1, 2, 3. When comparing models M3 and M4, it was evident that for the first sample realization $\mathbf{y_1}$, wherein significant deviations of the measured values from the reference values were observed in the negative part of the moderate scale, the RMSE values indicate that R_C_$M_3 > R_C$_M_4 and R_P_$M_3 < R_P$_M_4. For measurements $\mathbf{y_2}$ and $\mathbf{y_3}$, the reverse inequalities hold true. Additionally, it can be observed that for sample realization $\mathbf{y_2}$, the RMSE values R_P_M_3 and R_P_M_4 were significantly higher compared to the values for measurements $\mathbf{y_1}$ and $\mathbf{y_3}$. Despite having a broader tolerance interval across almost all points on the moderate scale, model M3 exhibited higher RMSE values for *total*_R_c, *total*_R_P, and *total*_p_c. For all measurements $\mathbf{y_i}$, i = 1, 2, 3, it holds that p_C_$M_1 < p_C$_M_2, and likewise, p_C_$M_3 < p_C$_M_4.

3.3. Comparison of Models Using Metrics Related to the Confusion Matrix

Recent studies have shown that risk assessment models can be compared using metrics relates to the confusion matrix [18,34]. Expressions for evaluating the global risk of producers and consumers from equations (28) and (31) are a kind of classifier that sorts measurements as falsely rejected (FR) and falsely accepted (FA), so that R_P = FR and R_C = FA. Considering that the measurements within the tolerance interval and the acceptance interval were defined as true positive (TP), and that the measurements outside the acceptance interval and tolerance interval were defined as true negative (TN), it is feasible to construct a confusion matrix [18]. For the confusion matrix constructed this way, especially, it is true that TP + TN + FR + FA=1. In addition, it holds that TP = $p_C - R_P$

and that $TN = 1 - p_C - R_C$ [18]. For a well-performed measurement, it is always the case that $TP \gg TN$, i.e., it is about imbalanced data [65]. Since, in the measurements there is always a measurement uncertainty, in practice, it is assumed that the risks R_P and R_C are always present and that their values differ from zero. Theoretically, in metrology, these values can be equal to zero, but not simultaneously.

The models outlined in the article can be evaluated by utilizing any of the commonly recognized metrics: accuracy, precision, recall, F1 score, Matthew's correlation coefficient (MCC), Cohen's Kappa, etc. [66,67]. A comparison of models by using metrics associated with confusion matrices was conducted for the area enclosed by the moderate scale and guard band axes. In this article, the F1 score, and diagnostic odds ratio (DOR) were chosen for model comparison. The F1 score is the harmonic mean of precision and recall. According to [18], in the metrological sense, it can be written in the form:

$$F1 = \frac{p_C - R_P}{p_C - R_P + \frac{R_P + R_C}{2}}. \tag{38}$$

This metric is conveniently used in binary classification when it is necessary to recognize and classify a specific class of confusion matrix [68]. Here, it is the TP class. Standard values of the F1 score range from $[0,\ 1]$. In the metrological sense, the values 0 and 1 should be excluded. If the value of the F1 score were to be zero, it would be true that $R_P = p_C$, thereby indicating there is no TP measurement. Also, if the value of the F1 score were to be one, it would mean that $R_C + R_P = 0$, which is impossible because of measurement uncertainty. This is the rationale behind why the values of the F1 score, in the realm of metrology, ought to be within the interval $\langle 0, 1 \rangle$. The model's performances are better for values of the F1 score that are closer to one. Furthermore, the curves of the F1 score along the guard band axis pass through the intersection of the precision and recall curves. That characteristic intersection point is obtained for that value of the guard band for which $R_C = R_P$ [18]. The lengths of the acceptance intervals are chosen precisely so that the intersection of the risk surfaces will be visible on the graphs (Figures S3–S6). Like the risk surfaces, the F1 score surfaces for models M1 and M2 behave differently compared to models M3 and M4 (Figure 14).

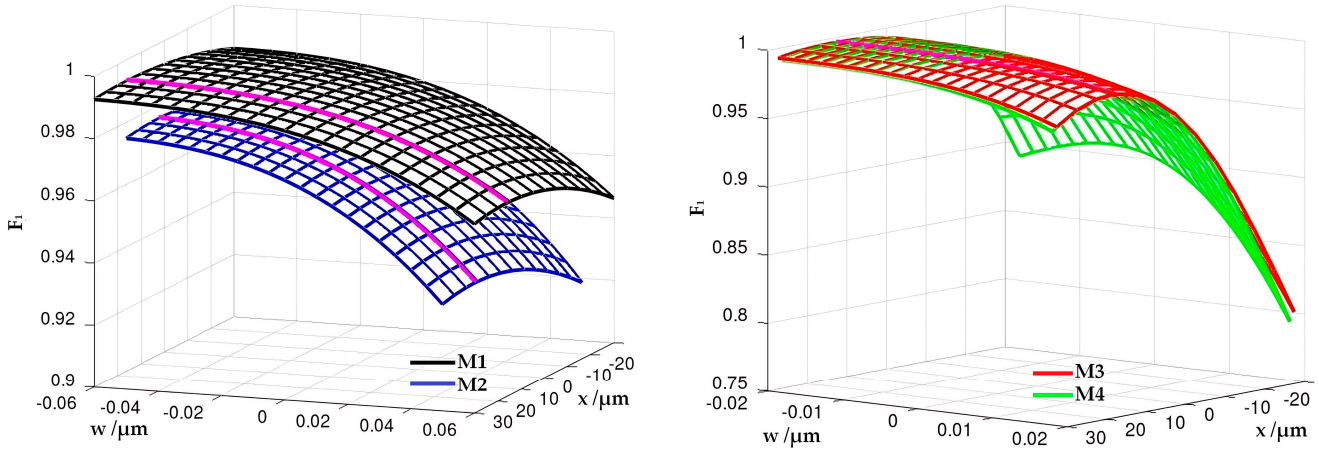


Figure 14. Comparison of risk models using F1 score: (**a**) F1 score for models M1 and M2; (**b**) F1 score for models M3 and M4. The curve of maximums along the guard band axis for all models is marked in pink.

For models M1 and M2, the F1 score decreases along the guard band axis. In contrast, the F1 score curves of models M3 and M4, along the moderate scale, have the maximum values at the points $(x_1, w_8), (x_2, w_7), \ldots, (x_8, w_1), \ldots, (x_{13}, w_1)$. By using the F1 score, it

can be seen that models M3 and M4 better detected deviations of the measured values from the reference values of the moderate scale on the edge portions of the guard band axis. For all models, the maximal values of the F1 score along the guard band axis are found at points $(x_{min}, w_k), k = 1, 2, \ldots, n_3$.

The comparison of risk assessment models was additionally performed using the DOR metric. DOR is the ratio of the probability of true positive measurements among falsely rejected measurements and the probability of falsely accepted measurements among true negative measurements [69]. It can be calculated from the following equation written in metrology terms:

$$DOR = \frac{(p_C - R_P)(1 - p_C - R_C)}{R_C R_P}. \quad (39)$$

Standard values for the DOR are in the range $[0, \infty\rangle$. If the value for DOR were equal to zero, it would mean that $R_C + R_P = 1$, that is, that $TP + TN = 0$. This would mean that the measurement was carried out disastrously. Therefore, in the metrological sense, the values of DOR must be in the interval $\langle 0, \infty\rangle$. Models that possess DOR values exceeding one are more effective in detecting probabilities of TP measurements among FR measurements [70]. The minimal value of DOR along the guard band axis is achieved for the shared risk model when the length of the guard band is $w_{11} = 0$. The highest values for DOR are achieved for low rates of the global consumer's and producer's risk [71]. Therefore, this metric is deemed suitable for identifying the line located in the plane that is enclosed by the moderate scale and by the guard band axis, along which the risks are the smallest. As can be seen from Figure 15a and Figure 15b, the curve that indicates the maximum values for the DOR is situated just above the line that traverses the points $(x_{min}, w_k), k = 1, 2, \ldots, n_3$. DOR for M3 and M4 models detects this curve extremely well (Figure 15b).

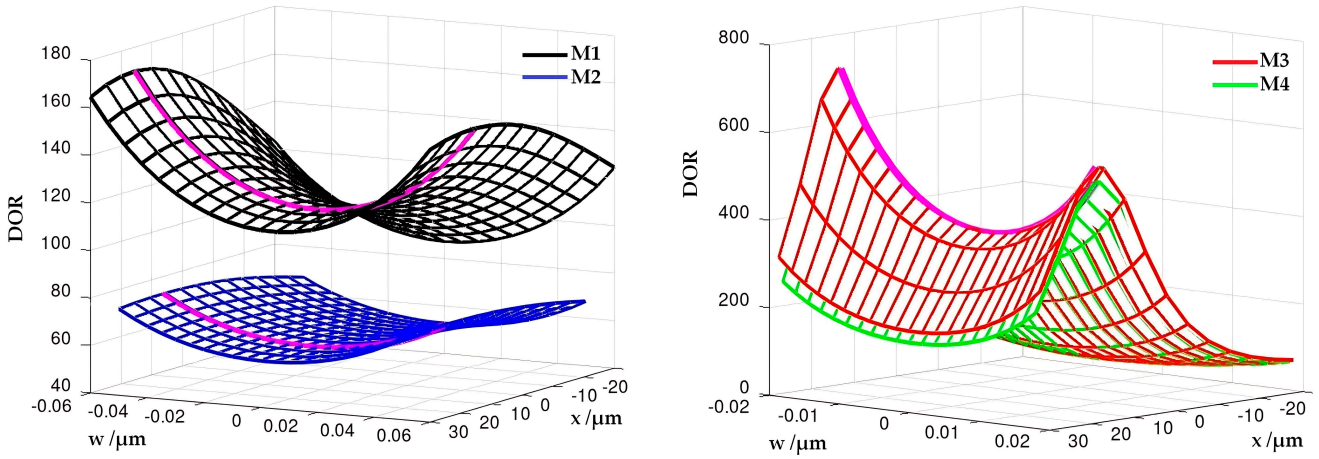
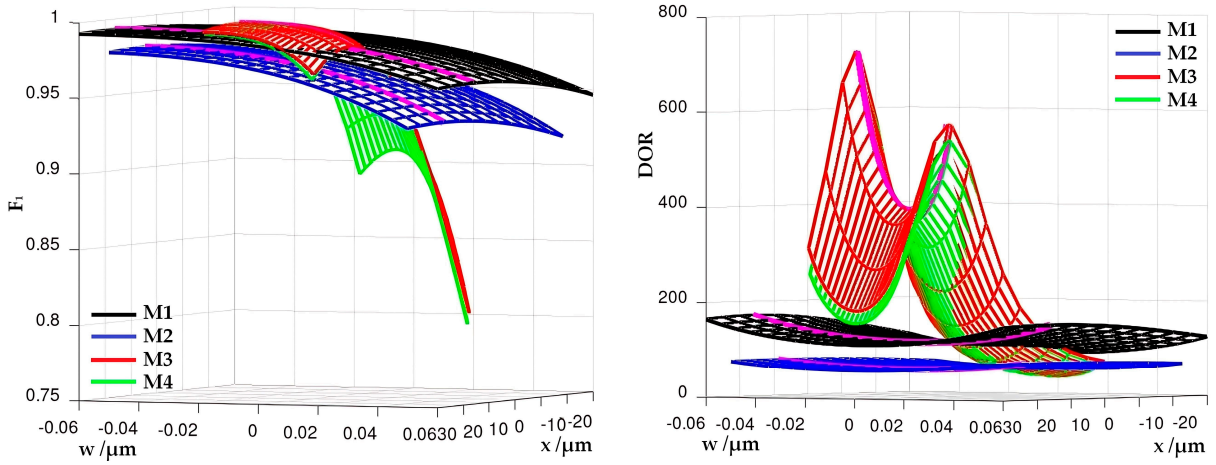


Figure 15. Comparison of the risk models using DOR: (**a**) DOR surfaces for models M1 and M2; (**b**) DOR surfaces for models M3 and M4. The curve of maximums for all models is marked in pink.

Similarly to the surfaces of the risks (Figure 9), there is an area where the F1 score of the models M3 and M4 has a higher value than the F1 score of the models M1 and M2 (Figure 16a). Analogously, there is an area where the DOR of models M3 and M4 has lower values compared to models M1 and M2 (Figure 16b).



(**a**) (**b**)

Figure 16. Comparison of models using F1 score and DOR surfaces: (**a**) F1 score; (**b**) DOR. The curve of maximums for all models is marked in pink.

4. Conclusions

The analysis indicates that it is acceptable for the iso-risk curves along the moderate scale to have a parabola shape with an upward opening. The minimal value of these parabolas is located at the point that is the intersection of the regression line and the $y = x$ line. The appearance of anomalies depends on the width of the tolerance interval ΔT and the value of the measurement uncertainty u_m of a future inspection process. These anomalies can be rectified by extending the tolerance interval range. Models, where for the risk calculation was used combined measurement uncertainty determined by applying a simplified measurement model with the use of comprehensive data on measurement performance, exhibit stable behavior along the moderate scale. Conversely, models in which the measurement uncertainty is calculated using the law of the propagation of uncertainty, applied to the functional relationship of the input data obtained from the LRA, are biased towards risks on the edge of the moderate scale. This was confirmed by testing the models using the F1 score. Furthermore, the RMSE for the mentioned models is better able to detect deviations of the measured values from the reference values on a moderate scale. DOR metrics possess an exceptional ability to detect the curve along the guard band axis, where the risks are the lowest. Therefore, in the risk assessment procedure for linear regression, it is advisable to use models where the measurement uncertainty is calculated by LRA.

The described method of risk assessment in regression models should be further investigated. Primarily, this refers to testing models for the different values of input parameters and testing their mutual relationships. Research can also focus on using the method in the case of polynomial and exponential regression models. The adaptation of the method and its application in calibration are also a matter of concern for future work. The method could be applied for risk assessment in models that use information from economics, insurance, medicine, or some other fields.

The risk assessment procedure for linear regression is a significant additional step in the analysis of the quality of the measurement. The application of this method certainly contributes to the improvement of product quality. Hence, risk assessment should be adopted as a standard procedure when assessing the conformity of products with given specifications, and in metrology in general.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app14062605/s1, Table S1: Shapiro–Wilk normality test results; Figure S1: Residual diagnostics; Table S2: Measurement uncertainties for each point of the moderate scale **x** calculated by LRA; Figure S2: The comparison of the RMSE values for conformance probability;

Figure S3: Risk surfaces for model M1; Figure S4: Risk surfaces for model M2; Figure S5: Risk surfaces for model M3; Figure S6: Risk surfaces for model M4.

Author Contributions: Conceptualization, D.B.; methodology, D.B. and B.R.; software, D.B.; validation, D.B., B.R. and A.R.; formal analysis, D.B. and A.R.; investigation, D.B.; resources, D.B., B.R. and A.R.; data curation, D.B. and B.R.; writing—original draft preparation, D.B.; writing—review and editing, B.R. and A.R.; visualization, D.B.; supervision, B.R.; project administration, A.R.; funding acquisition, B.R., D.B. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Savković, B.; Kovač, P.; Rodić, D.; Štrbac, B.; Klančnik, S. Comparison of artificial neural network, fuzzy logic, and genetic algorithm for cutting temperature and surface roughness prediction during the face milling process. *Adv. Prod. Eng. Manag.* **2020**, *15*, 137–150. [CrossRef]
2. Sheth, S.; Modi, B.; Patel, D.; Chaudhari, A. Modeling and Prediction Using Regression, ANN, and Fuzzy Logic of Real Time Vibration Monitoring on Lathe Machine in Context of Machining Parameters. *Bonfring. Int. J. Man. Mach. Interface* **2015**, *3*, 30–35. [CrossRef]
3. Razumić, A.; Runje, B.; Lisjak, D.; Kolar, D.; Horvatić Novak, A.; Štrbac, B.; Savković, B. Atomic force microscopy: Step height measurement uncertainty evaluation. In Proceedings of the International Conference MATRIB 2023—Materials, Tribology & Recycling, Vela Luka, Croatia, 29 June–1 July 2023.
4. Papafotis, K.; Nikitas, D.; Sotiriadis, P.P. Magnetic Field Sensors' Calibration: Algorithms' Overview and Comparison. *Sensors* **2021**, *21*, 5288. [CrossRef] [PubMed]
5. Shen, G.; Wang, Y.; Dewaele, A.; Wu, C.; Fratanduono, D.E.; Eggert, J.; Klotz, S.; Dziubek, K.F.; Loubeyre, P.; Fat'yanov, O.V.; et al. Toward an international practical pressure scale: A proposal for an IPPS ruby gauge (IPPS-Ruby2020). *High Press. Res.* **2020**, *40*, 299–314. [CrossRef]
6. Aime, L.F.J.; Kissinger, T.; James, S.W.; Chehura, E.; Verzeletti, A.; Tatam, R.P. High sensitivity pressure measurement using optical fibre sensors mounted on a composite diaphragm. *Opt. Express.* **2021**, *29*, 4105–4123. [CrossRef] [PubMed]
7. Greaves, M.; Caillon, N.; Rebaubier, H.; Bartoli, G.; Bohaty, S.; Cacho, I.; Clarke, L.; Cooper, M.; Daunt, C.; Delaney, M.; et al. Interlaboratory comparison study of calibration standards for foraminiferal Mg/Ca thermometry. *Geochem. Geophys. Geosyst.* **2008**, *9*, Q08010. [CrossRef]
8. Velychko, O.; Shevkun, S.; Gordiyenko, T.; Mescheriak, O. Interlaboratory comparisons of the calibration results of time meters. *East. Eur. J. Enterp. Technol.* **2018**, *1*, 4–11. [CrossRef]
9. Wübbeler, G.; Bodnar, O.; Elster, C. Robust Bayesian linear regression with application to an analysis of the CODATA values for the Planck constant. *Metrologia* **2018**, *55*, 20. [CrossRef]
10. Liao, K.; Shafieloo, A.; Keeley, R.E.; Linder, E.V. A Model-independent Determination of the Hubble Constant from Lensed Quasars and Supernovae Using Gaussian Process Regression. *Astrophys. J. Lett.* **2019**, *886*, L23. [CrossRef]
11. Cox, M.G.; Forbes, A.B.; Harris, P.M.; Smith, I.M. *The Classification and Solution of Regression Problems for Calibration, NPL Report CMSC 24/03*; National Physical Laboratory: Teddington, UK, 2004. Available online: https://eprintspublications.npl.co.uk/2772/1/cmsc24.pdf (accessed on 4 January 2024).
12. Fernández-Delgado, M.; Sirsat, M.S.; Cernadas, E.; Alawadi, S.; Barro, S.; Febrero-Bande, M. An extensive experimental survey of regression methods. *Neural Netw.* **2019**, *111*, 11–34. [CrossRef]
13. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. Evaluation of Measurement Data—The Role of Measurement Uncertainty in Conformity Assessment. Joint Committee for Guides in Metrology, JCGM 106:2012. BIPM. 2012. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_106_2012_E.pdf/fe9537d2-e7d7-e146-5abb-2649c3450b25 (accessed on 4 January 2024).
14. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. Evaluation of Measurement Data—An Introduction to the "Guide to the Expression of Uncertainty in Measurement" and Related Documents. Joint Committee for Guides in Metrology, JCGM 104:2009. BIPM. 2009. Available online: https://www.bipm.org/en/committees/jc/jcgm/publications (accessed on 4 January 2024).
15. Expression of uncertainty in measurement. *Chem. Int.* **2018**, *40*, 30–31. [CrossRef]
16. GUM-Introduction. Available online: https://www.iso.org/sites/JCGM/GUM-introduction.htm (accessed on 4 January 2024).

17. *ILAC-G8:09/2019*; Guidelines on Decision Rules and Statements of Conformity. ILAC Secretariat: Newton, SA, Australia, 2019. Available online: https://ilac.org/publications-and-resources/ilac-guidance-series/ (accessed on 5 January 2024).
18. Božić, D.; Runje, B.; Lisjak, D.; Kolar, D. Metrics Related to Confusion Matrix as Tools for Conformity Assessment Decisions. *Appl. Sci.* **2023**, *13*, 8187. [CrossRef]
19. Pendrill, L.R. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia* **2014**, *51*, 3206. [CrossRef]
20. Dias, F.R.S.; Lourenço, F.R. Measurement uncertainty evaluation and risk of false conformity assessment for microbial enumeration tests. *J. Microbiol. Methods* **2021**, *189*, 106312. [CrossRef]
21. Williams, A.; Magnusson, B. (Eds.) Eurachem/CITAC Guide: Use of Uncertainty Information in Compliance Assessment. Available online: https://www.eurachem.org/images/stories/Guides/pdf/MUC2021_P1_EN.pdf (accessed on 9 January 2024).
22. Young, D.S. Tolerance: An R Package for Estimating Tolerance Intervals. *J. Stat. Softw.* **2010**, *36*, 1–39. [CrossRef]
23. Wallis, A.W. Tolerance intervals for linear regressions. In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 31 July–12 August 1950. Available online: https://digicoll.lib.berkeley.edu/record/112745/files/math_s2_article-04.pdf (accessed on 5 January 2024).
24. EUROLAB. Technical Report No.1/2017-Decision Rules Applied to Conformity Assessment. Available online: https://www.eurolab.org/pubs-techreports (accessed on 7 January 2024).
25. Božić, D.; Samardžija, M.; Kurtela, M.; Keran, Z.; Runje, B. Risk Evaluation for Coating Thickness Conformity Assessment. *Materials* **2023**, *16*, 758. [CrossRef]
26. Runje, B.; Horvatić Novak, A.; Razumić, A.; Piljek, P.; Štrbac, B.; Orošnjak, M. Evaluation of Consumer and Producer Risk in Conformity Assessment Decision. In Proceedings of the 30th DAAAM International Symposium "Intelligent Manufacturing & Automation", Zadar, Croatia, 23–26 October 2019. [CrossRef]
27. Božić, D.; Runje, B. Data Modelling in Risk Assessment. In Proceedings of the Laboratory Competence-2022, Cavtat, Croatia, 9–12 November 2022. Available online: https://www.crolab.hr/userfiles/file/cavtat2022/CROLAB_Cavtat%202022_zbornik%20radova_final_B.pdf (accessed on 9 January 2024).
28. Toczek, W.; Smulko, J. Risk Analysis by a Probabilistic Model of the Measurement Process. *Sensors* **2021**, *21*, 2053. [CrossRef]
29. Rajan, A.; Kuang, Y.C.; Po-Leen Ooi, M.; Demidenko, S.N. Moments and Maximum Entropy Method for Expanded Uncertainty Estimation in Measurements. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Turin, Italy, 22–25 May 2017. [CrossRef]
30. Weise, K.; Woger, W. A Bayesian theory of measurement uncertainty. *Meas. Sci. Technol.* **1993**, *4*, 1. [CrossRef]
31. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method. Joint Committee for Guides in Metrology, JCGM 101:2008. BIMP. 2008. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c (accessed on 9 January 2024).
32. Lira, I. A Bayesian approach to the consumer's and producer's risks in measurement. *Metrologia* **1999**, *36*, 397–402. [CrossRef]
33. Cox, M.G.; Forbes, A.B.; Harris, P.M. Bayesian estimation methods in metrology. In Proceedings of the 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Garching, Germany, 25–30 July 2004. [CrossRef]
34. Božić, D.; Runje, B. Selection of an Appropriate Prior Distribution in Risk Assessment. In Proceedings of the 33rd International DAAAM Virtual Symposium "Intelligent Manufacturing & Automation", Vienna, Austria, 26–27 October 2022. [CrossRef]
35. Brandão, L.P.; Silva, V.F.; Bassi, M.; de Oliveira, E.C. Risk Assessment in Monitoring of Water Analysis of a Brazilian River. *Molecules* **2022**, *27*, 3628. [CrossRef]
36. Kuselman, I.; Pennecchi, F.; Bettencourt da Silva, R.J.N.; Brynn Hibbert, D. Conformity assessment of multicomponent materials or objects: Risk of false decisions due to measurement uncertainty—A case study of denatured alcohols. *Talanta* **2017**, *164*, 189–195. [CrossRef]
37. Pennecchi, F.R.; Kuselman, I.; Hibbert, B.D.; Sega, M.; Rolle, F.; Altshul, V. Fit-for-purpose risks in conformity assessment of a substance or material—A case study of synthetic air. *Measurement* **2022**, *188*, 110542. [CrossRef]
38. Pennecchi, F.R.; Kuselman, I.; Di Rocco, A.; Brynn Hibbert, D.; Semenova, A.A. Risks in a sausage conformity assessment due to measurement uncertainty, correlation, and mass balance constraint. *Food Control* **2021**, *125*, 107949. [CrossRef]
39. Separovic, L.; de Godoy Bertanha, M.L.; Saviano, A.M.; Lourenço, F.R. Conformity Decisions Based on Measurement Uncertainty—A Case Study Applied to Agar Diffusion Microbiological Assay. *J. Pharm. Innov.* **2020**, *15*, 110–115. [CrossRef]
40. Lombardo, M.; Margueiro da Silva, S.; Lourenço, F.R. Conformity assessment of medicines containing antibiotics—A multi-variate assessment. *Regul. Toxicol. Pharmacol.* **2022**, *136*, 105279. [CrossRef]
41. Pennecchi, F.R.; Kuselman, I.; Bettencourt da Silva, R.J.N.; Brynn Hibbert, D. Risk of a false decision on conformity of an environmental compartment due to measurement uncertainty of concentrations of two or more pollutants. *Chemosphere* **2018**, *202*, 165–176. [CrossRef]
42. Separovic, L.; Lourenço, F.R. Measurement uncertainty and risk of false conformity decision in the performance evaluation of liquid chromatography analytical procedures. *J. Pharm. Biomed. Anal.* **2019**, *171*, 73–80. [CrossRef]
43. Caffaro, A.M.; Lourenço, F.R. Total combined global risk assessment applied to pharmaceutical equivalence—A case study of ofloxacin medicines. *Chemom. Intell. Lab. Syst.* **2023**, *241*, 104935. [CrossRef]

44. Bednjanec, F. Umjeravanje Uređaja za Mjerenje Kružnosti. Diplomski rad, Fakultet Strojarstva i Brodogradnje, Sveučilište u Zagrebu 2016. 24 March 2016. Available online: https://urn.nsk.hr/urn:nbn:hr:235:701539 (accessed on 10 January 2024).
45. Huzak, M. *Vjerojatnost i Matematička Statistika, Predavanja*; Poslijediplomski Specijalistički Sveučilišni Studij Aktuarske Matematike; Specialist u Zagrebu, PMF-Matematički Odjel: Zagreb, Croatia, 2006. Available online: http://aktuari.math.pmf.unizg.hr/docs/vms.pdf (accessed on 11 January 2024).
46. Ortiz, M.C.; Sánchez, M.S.; Sarabia, L.A. 1.05—Quality of Analytical Measurements: Univariate Regression. In *Comprehensive Chemometrics. Chemical and Biochemical Data Analysis*, 1st ed.; Brown, S.D., Tauler, R., Walczak, B., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; Volume 1, pp. 127–169. [CrossRef]
47. Ellison, S.R.; Williams, A. (Eds.) Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement. Available online: https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf (accessed on 11 January 2024).
48. Miller, S.J. *The Method of Least Squares*; Mathematics Department Brown University: Providence, RI, USA, 2006. Available online: https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/MethodLeastSquares.pdf (accessed on 15 January 2024).
49. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. Joint Committee for Guides in Metrology, JCGM 100:2008. BIPM. 2008. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6 (accessed on 16 January 2024).
50. Taylor, B.N.; Kuyatt, C.E. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*; NIST Technical Note 1297; US Department of Commerce, Technology Administration, National Institute of Standards and Technology: Gaithersburg, MD, USA, 1994. Available online: https://emtoolbox.nist.gov/publications/nisttechnicalnote1297s.pdf (accessed on 16 January 2024).
51. Farrance, I.; Frenkel, R. Uncertainty of Measurement: A Review of the Rules for Calculating Uncertainty Components through Functional Relationships. *Clin. Biochem. Rev.* **2012**, *33*, 49–75. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387884/ (accessed on 16 January 2024).
52. Klauenberg, K.; Martens, S.; Bošnjaković, A.; Cox, M.G.; Van der Veen, A.M.; Elster, C. The GUM perspective on straight-line errors-in-variables regression. *Measurement* **2021**, *187*, 110340. [CrossRef]
53. Croarkin, C.; Tobias, P.; Zey, C. *Engineering Statistics Handbook*; The Institute Gaithersburg: Gaithersburg, MD, USA, 2001. Available online: https://www.itl.nist.gov/div898/handbook/dtoc.htm (accessed on 16 January 2024).
54. Krishnamoorthy, K.; Mathew, T. *Statistical Tolerance Regions: Theory, Applications, and Computation*; Wiley: Hoboken, NJ, USA, 2009.
55. Splinter, K.; Sigler, G.; Harman, M.; Kolsti, K. *Tolerance Intervals Demystified*; STAT Center of Excellence Air Force Institute of Technology: Wright-Patterson AFB, OH, USA. Available online: https://www.afit.edu/STAT/statcoe_files/Tolerance%20Intervals%20Demystified.pdf (accessed on 23 January 2024).
56. Greenwell, B.M. Topics in Statistical Calibration. Ph.D. Thesis, Air Force Institute of Technology, Hobson Way, OH, USA, 3 March 2014. Available online: https://apps.dtic.mil/sti/pdfs/ADA598921.pdf (accessed on 23 January 2024).
57. R Core Team. *R: A Language and Environment for Statistical Computing*; The R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed on 13 February 2024).
58. Borchers, H.W. Pracma: Practical Numerical Math Functions. R Package Version 2.4.2/r532. Available online: https://R-Forge.R-project.org/projects/optimist/ (accessed on 13 February 2024).
59. Eaton, J.W.; Bateman, D.; Hauberg, S.; Wehbring, R. GNU Octave Version 8.4.0 Manual: A High-Level Interactive Language for Numerical Computations. Available online: https://www.gnu.org/software/octave/doc/v8.4.0/ (accessed on 13 February 2024).
60. Botchkarev, A. A new typology design of performance metrics to measure errors in machine learning regression algorithms. *Interdiscip. J. Inf. Knowl. Manag.* **2019**, *14*, 45–76. [CrossRef]
61. Abhishek, T. Comparative Assessment of Regression Models Based on Model Evaluation Metrics. *Int. Res. J. Eng. Technol.* **2021**, *9*, 853–860.
62. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]
63. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [CrossRef] [PubMed]
64. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]
65. Jeni, L.A.; Cohn, J.F.; de la Torre, F. Facing Imbalanced Data—Recommendations for the Use of Performance Metrics. In Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), Geneva, Switzerland, 2–5 September 2013. [CrossRef]
66. De Diego, I.M.; Redondo, A.R.; Fernández, R.R.; Navarro, J.; Moguerza, J.M. General Performance Score for Classification Problems. *Appl. Intell.* **2022**, *52*, 12049–12063. [CrossRef]
67. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756. [CrossRef]
68. Chicco, D.; Tötsch, N.; Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* **2021**, *14*, 13. [CrossRef]

69. McHugh, M.L. The odds ratio: Calculation, usage, and interpretation. *Biochem. Medica* **2009**, *19*, 120–126. Available online: https://hrcak.srce.hr/37593 (accessed on 1 February 2024). [CrossRef]
70. Glas, A.S.; Lijmer, J.G.; Prins, M.H.; Bonsel, G.J.; Bossuyt, P.M.M. The diagnostic odds ratio: A single indicator of test performance. *J. Clin. Epidemiol.* **2003**, *56*, 1129–1135. [CrossRef]
71. Šimundić, A.M. Measures of Diagnostic Accuracy: Basic Definitions. *Ejifcc* **2009**, *19*, 203. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975285/ (accessed on 1 February 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.