Review

A Review of Methods for Assessing the Quality of Measurement Systems and Results

Andrej Razumić [1], Biserka Runje [2], Vesna Alar [2,*], Branko Štrbac [3] and Zvonko Trzun [1]

[1] Department of Polytechnics, Dr. Franjo Tuđman Defense and Security University, Ilica 256b, 10000 Zagreb, Croatia; andrej.razumic@sois-ft.hr (A.R.); zvonko.trzun@sois-ft.hr (Z.T.)

[2] Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lučića 5, 10000 Zagreb, Croatia; biserka.runje@fsb.unizg.hr

[3] Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia; strbacb@uns.ac.rs

* Correspondence: vesna.alar@fsb.unizg.hr

Abstract

This paper provides an overview of key methods for assessing the quality of measurement systems and measurement results. Fundamental concepts are analyzed and their definitions compared across various metrological standards. Particular emphasis is placed on the issue of terminological inconsistency, as the same terms are interpreted differently in different documents, which may lead to misinterpretation or incorrect application of methods. This paper also presents practical examples of how these methods are applied to real data. Clear indications are provided regarding the conditions under which each method is recommended, depending on factors such as model complexity, data availability, and intended application. In addition, it offers a discussion and recommendations for future directions, highlighting the need for harmonized terminology, standardization of evaluation procedures, and the adoption of advanced technologies such as artificial intelligence and machine learning in the assessment of measurement quality.

Keywords: measurement system analysis; measurement uncertainty; metrological standards; terminological inconsistency

1. Introduction

In industrial and scientific environments, where decisions, control processes, and analyses heavily rely on measurement data, ensuring the quality of measurement results, methods, and systems is a fundamental prerequisite for reliability, safety, and compliance with technical requirements. Increasing product complexity, tighter regulatory criteria, and the advancement of digitally controlled production and diagnostic processes further emphasize the need for accurate, traceable, and reliable measurements [1,2].

To achieve a credible assessment of measurement quality and thereby ensure reliable measurement results, a wide array of methods have been developed, ranging from practical tools used in industry to scientifically grounded approaches for evaluating measurement uncertainty. Over the last few decades, various normative documents, guides, and methodological frameworks have been established for this purpose. Among the most prominent are the industry-oriented Measurement System Analysis (MSA) [3]; the ISO 5725 standard series on Accuracy (trueness and precision) of measurement methods and results (Parts 1–6) [4–9]; and the Guide to the Expression of Uncertainty in Measurement

(GUM, JCGM 100:2008) [10], which focuses on the quantitative expression of uncertainty based on mathematical modeling of all significant sources.

In addition to traditional uncertainty propagation approaches, Monte Carlo simulations as addressed in JCGM 101:2008 [11] and Bayesian method are increasingly applied, particularly in complex measurement scenarios involving nonlinear relationships or limited data [12–14]. Furthermore, ISO 21748:2017 [15] provides guidance on evaluating measurement uncertainty using repeatability, reproducibility, and trueness data derived from interlaboratory studies conducted in accordance with ISO 5725-2 [5].

The evaluation of measurement uncertainty is a challenging task, particularly for complex measurement systems, and represents one of the most prominent research directions in modern metrology [16,17]. It is essential for maintaining metrological traceability and for ensuring conformity with specifications [18]. A measurement result is considered incomplete if it does not include an interval describing the potential dispersion of values that is, its uncertainty. The true value of the measurand is expected to lie within this uncertainty interval with a defined probability. For this reason, the evaluation of measurement uncertainty is a mandatory requirement for accredited calibration and testing laboratories performing calibrations of standards, instruments, and measurement artifacts [19].

GUM is the globally recognized standard for quantifying uncertainty in measurement results. It is based on mathematical modeling of Type A and Type B sources of uncertainty, followed by analytical uncertainty propagation. Although highly robust, GUM requires a deep understanding of statistical models, detailed knowledge of input variables, and often the use of specialized computational tools, including those for Monte Carlo simulations or Bayesian inference. In cases where analytical models are not feasible, Monte Carlo methods (JCGM 101) or Bayesian approaches may be employed to enable uncertainty evaluation in data-limited or nonlinear systems [20,21].

In industrial practice, however, the evaluation of measurement uncertainty is often avoided or simplified for practical, organizational, or methodological reasons [22]. In mass production environments, MSA is frequently used to assess the percentage contribution of the measurement system to total data variability, the number of distinguishable categories, and the repeatability and reproducibility of results. This approach avoids the complexity of full uncertainty modeling and instead focuses on assessing measurement system capability [23]. In many industrial settings, there is a shortage of trained metrology professionals and limited time or resources for full uncertainty evaluations, especially for routine measurements. Additionally, production environments often prioritize process efficiency and compliance with tolerance limits rather than detailed quantification of each uncertainty component. As a result, measurement uncertainty is sometimes viewed as bureaucratic overhead, unless explicitly required by accreditation, certification, or regulatory authorities.

Many engineers mistakenly believe that Gage Repeatability and Reproducibility (GR&R) "evaluates uncertainty", which is not accurate. GR&R analyzes only the variability introduced by the measurement system and does not provide information on the interval within which the true value of the measurand is expected to lie [23,24].

The ISO 5725 standard series was developed to rigorously quantify the accuracy, trueness, and precision of measurement methods. It is widely used in laboratory and research settings, particularly in accredited testing laboratories under ISO/IEC 17025 [25] and in highly regulated industries such as pharmaceuticals, food, and chemicals [26,27]. ISO 5725 is also a foundation for evaluating repeatability and reproducibility, conducting interlaboratory comparisons. When used in combination with ISO 21748, it provides an experimental foundation for evaluating measurement uncertainty [28]. However, in day-to-day industrial settings, its application is less common due to its complexity, need for comprehensive statistical testing, and the requirement for extensive experimental design.

Moreover, the standard uses precise distinctions among terms such as trueness, accuracy, and bias, terms that are often misunderstood or used inconsistently in practice.

Despite their shared objective to quantify and clearly express measurement quality, the reviewed approaches differ in terminology, modeling, and evaluation criteria. A notable challenge is the inconsistent use of fundamental metrological terms (e.g., accuracy, precision, repeatability, reproducibility, trueness, and bias) across standards and industries. This terminological fragmentation impedes communication among metrologists, quality engineers, and technical professionals.

Although various studies focus on specific aspects of GUM, MSA, or ISO 5725, comprehensive comparative reviews remain scarce, particularly those that bridge theoretical concepts with practical implementations.

This review adopts the premise that a comparative and integrated understanding of these methodologies can enhance the selection of appropriate approaches across different metrological and industrial contexts and support their integration in modern, digitally assisted measurement systems.

The objectives of this review paper are to

(i)   compare how key metrological concepts (accuracy, precision, repeatability, reproducibility, trueness, bias, and measurement uncertainty) are defined and applied in major frameworks (MSA, ISO 5725, GUM, ISO 21748, etc.);
(ii)   illustrate their application through real or literature-based data examples;
(iii)   identify key terminological inconsistencies that may lead to misinterpretation;
(iv)   offer guidelines for method selection and harmonization, including the integration of digital and AI-based tools.

2. Comparison of Terminology, Standards, and Methodologies

This section presents an overview of the fundamental terminology and concepts as defined in the International Vocabulary of Metrology (VIM), ISO 5725, GUM, and MSA Reference Manual, with emphasis on their similarities, differences, and relevance within the context of this paper. A comparison of the definitions of key metrological concepts is presented in Table 1. This provides the terminological foundation for comparing the methodologies discussed later in this section.

In order to harmonize and update metrological terminology, international standards and guides are continuously revised to reflect the latest scientific and practical insights. One of the key documents in this domain, the International Vocabulary of Metrology (VIM), is currently under revision. Since January 2021, the fourth edition has been in preparation (VIM4 CD—Committee Draft), introducing important terminological and conceptual clarifications regarding fundamental metrological concepts such as trueness, accuracy, and measurement error. The aim of this revision is to achieve better alignment with various standards (including ISO 5725 and the GUM), enhance clarity in the interpretation of measurement results, and broaden the applicability of definitions across all scientific and engineering fields.

Although the document is currently in draft form and not intended for public cita-tion, its relevance should be emphasized, as it represents an important step toward future harmonization of metrology with the principles of best practice and the consistent understanding of metrological terms worldwide.

Table 1. Comparison of definitions of fundamental metrological concepts.

| Concept	Definition VIM	Definition GUM (JCGM 100)	Definition ISO 5725:2023	Definition MSA (Reference Manual Fourth Edition)
Measurement result	Set of quantity values attributed to a measurand along with relevant information.	An estimate of the value of a measurand accompanied by a statement of uncertainty.	Value obtained by applying a measurement procedure according to a test method.	Result of the measurement process involving the system, operator, and environment.
Measurement error	Difference between a measured quantity value and a reference quantity value; includes systematic and random components.	Not directly defined; discussed in the context of uncertainty. Error is a concept underlying uncertainty.	Overall error defined as a combination of trueness and precision.	Difference between the average result and the reference value (bias).
Precision	Closeness of agreement between indications or measured values obtained under specified conditions.	Quantified by standard deviation of measured values under repeatability or reproducibility conditions.	Based on the standard deviation in results under defined conditions.	Variability expressed as Equipment Variation (EV).
Trueness	Closeness of the mean of measured values to the reference value; inversely related to systematic error.	Discussed indirectly; related to the concept of systematic error.	Closeness of the mean result to the accepted reference value.	Not directly used; focus is on bias as an expression of trueness.
Accuracy	Combination of trueness and precision; not a quantity and not expressed numerically.	Sometimes used informally to mean closeness to true value, but not defined as a quantity.	Accuracy = trueness + precision. Quantified experimentally.	Often incorrectly used as a synonym for bias.
Bias	Estimate of a systematic measurement error, i.e., the difference between the expectation of test results and a reference value.	Systematic effect on a measurement result, often corrected or accounted for through uncertainty.	Difference between the expected result and the reference value.	Systematic error expressed as deviation from the reference value.
Repeatability and Reproducibility	Repeatability: same conditions. Reproducibility: different conditions across labs/operators.	Used as input in uncertainty models; considered under standard uncertainty estimates.	Repeatability: same conditions. Reproducibility: different labs, equipment, and operators.	EV (repeatability) and AV (reproducibility) via Gage R&R studies.
Measurement uncertainty	Non-negative parameter characterizing the dispersion of quantity values attributed to a measurand.	Parameter associated with the result of a measurement, characterizing dispersion of values that could be reasonably attributed to the measurand (uncertainty means doubt, and therefore in its broadest sense, uncertainty of measurement means doubt about the validity of the result of a measurement).	Is not explicitly defined (but it is clearly implied that repeatability, reproducibility, and trueness constitute experimental input data that can be used in the evaluation of measurement uncertainty, for example, in accordance with the GUM or ISO 21748, and bias (as an expression of trueness) should be corrected for if known, and the remaining uncertainty of the correction can be included in the overall measurement uncertainty).	Estimate of the range within which the true value is believed to lie.

Measurement System Analysis (MSA) originates from industrial quality control practices and is designed to evaluate the practical performance of a measurement system [29]. It focuses on identifying and quantifying sources of variability related to the instrument, operator, method, and environment. This is typically achieved through tools such as Gage R&R studies, bias analysis, linearity, and stability assessments [30–32]. The primary purpose of MSA is to ensure that the measurement system is capable of reliably distinguishing between conforming and non-conforming parts or outcomes [33,34]. MSA plays a central role in industrial manufacturing and process control particularly in sectors such as automotive, aerospace, and pharmaceuticals where precise and consistent measurement is critical [35,36]. It is widely used during gage validation, process qualification, and quality improvement initiatives, providing a practical framework for evaluating the variability of

a measurement system, with particular emphasis on repeatability, reproducibility, bias, and long-term stability. MSA methods, purpose, data type, recommended illustrations, and acceptance criteria are presented in Table 2.

Table 2. MSA methods, purpose, data type, recommended illustrations, and acceptance criteria.

MSA Method	Data Type	Purpose	Recommended Illustration	Acceptance Criteria
Gage R&R (Average & Range)	Variable	Basic assessment of repeatability and reproducibility of the measurement system	Scatter plot by parts and appraisers	GRR < 10%: acceptable, 10–30%: marginal, >30%: unacceptable
Gage R&R (ANOVA)	Variable	Detailed statistical analysis of variability components	Variance components chart, interaction plot	Same as above but includes significance tests for sources of variation
Gage R&R (Nested)	Variable	Used when the same parts are not available for all appraisers (e.g., destructive testing)	Box plot by appraiser	Interpret based on % contribution of nested sources; GRR guidelines apply
Expanded Gage R&R	Variable	Extended analysis including additional factors (e.g., time, device, and location)	Multifactor plot, interaction diagram	Acceptability depends on total variation explained; no fixed % thresholds
Type 1 Gage Study	Variable	Quick check of bias and repeatability of a single device	Histogram and $\overline{X}$ control chart	$Cg \geq 1.33$; $Cgk \geq 1.33$ (capability indices)
Linearity Study	Variable	Assessment of measurement bias across the measurement range	Scatter plot with regression line	Linearity slope ≈ 0; $R^2 > 0.9$; individual biases within limits
Bias Study	Variable	Assessment of deviation from known reference value	Bar chart (bias)	Bias within tolerance limits or statistically non-significant
Stability Study	Variable	Check of measurement system consistency over time	$\overline{X}$ and R control chart	No significant trend; all points within control limits
Attribute Agreement Analysis	Attribute	Assessment of rater consistency and agreement with standard	Accuracy matrix, agreement table	$\geq 90\%$ agreement recommended; $\geq 70\%$ for marginal acceptability
Kappa Analysis (Cohen/Fleiss)	Attribute	Statistical measure of inter-rater agreement (multiple raters)	Kappa heatmap	Kappa > 0.75: strong agreement; 0.4–0.75: moderate; <0.4: weak

Figure 1 illustrates the structure and key components of the MSA method applied in GR&R studies.
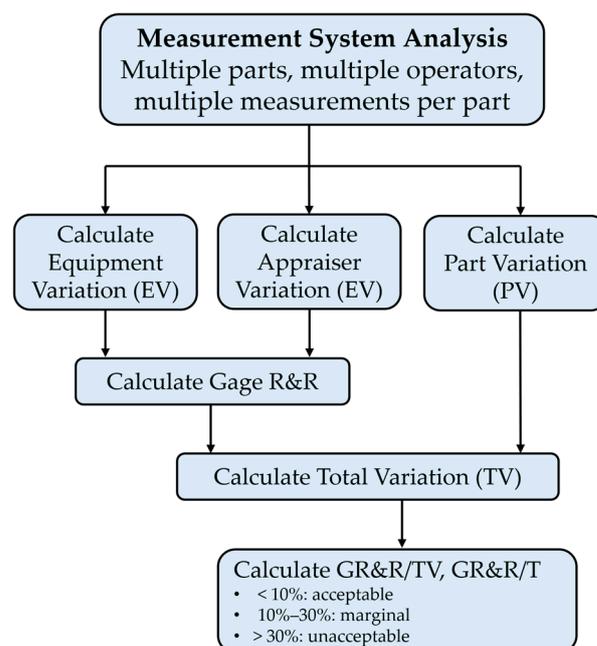
Figure 1. Key components of the MSA method used in GR&R analysis.

ISO 5725 is based on the experimental assessment of measurement method performance, with accuracy defined as a combination of trueness (closeness to a reference value) and precision (closeness of repeated results under specified conditions). This standard is widely used for method validation, interlaboratory comparisons, and estimating variability under repeatability and reproducibility conditions [37–39].

An overview of the ISO 5725 Series' purpose, application, recommended illustrations, and acceptance criteria is presented in Table 3.

Table 3. Overview of ISO 5725 series—purpose, application, recommended illustrations, and acceptance criteria.

Part	Purpose	Typical Application	Recommended Illustration	Acceptance Criteria
ISO 5725-1:2004	Introduces general principles for evaluating the accuracy of measurement methods and results, including definitions and framework.	Reference for terminology, planning accuracy studies, and understanding components of accuracy.	Concept diagrams showing relationship between accuracy, trueness, and precision.	No specific criteria; foundational concepts and definitions.
ISO 5725-2:2020	Provides basic methods for estimating repeatability and reproducibility through interlaboratory studies.	Evaluation of precision in collaborative tests; standard design for reproducibility studies.	Standard deviation charts, Youden plots, repeatability vs. reproducibility plots.	Repeatability and reproducibility standard deviations should be within limits relevant to the method or product specification.
ISO 5725-3:2024	Describes designs for obtaining intermediate precision and provides alternatives to designs in Part 2.	Used when variability due to operator, time, or equipment must be evaluated within the same lab.	Box plots, control charts stratified by operator or day.	Intermediate precision variability should be assessed in context of long-term method performance.
ISO 5725-4:2020	Provides basic methods for the determination of trueness using reference values.	Used to assess bias in measurement methods through comparison with reference materials.	Bias plots, difference plots, histograms comparing measurement and reference values.	Bias should be statistically insignificant or within acceptable limits of the method.
ISO 5725-5:2004/corr2009	Presents alternative methods for determining trueness and precision under specific conditions.	Applied in cases with limited data, non-standard conditions, or specific method constraints.	Simulation plots, adaptive model visualizations.	Acceptance depends on context-specific statistical validation; no fixed universal limits.
ISO 5725-6:2004	Provides practical examples of applications of the trueness and precision concepts.	Training, guidance for implementation, and case studies of statistical evaluations.	Applied case study charts, visual summaries of study outcomes.	Illustrative only; no acceptance criteria defined.

Complementing the summary in Table 3, Figure 2 illustrates the ISO 5725 framework for evaluating the accuracy of measurement methods, including both trueness and precision.

The measurement uncertainty evaluation based on the Guide to the Expression of Uncertainty in Measurement (GUM) follows a systematic and transparent approach to modeling the measurement process [17]. The first step is to define a measurement model, i.e., a mathematical equation that expresses the measurand as a function of input quantities that influence the result.

Each input quantity is then assigned a standard uncertainty, which is evaluated using Type A methods (statistical analysis of repeated measurements) or Type B methods (based on other information such as calibration certificates, manufacturer specifications, or expert judgment) [40]. These uncertainties are propagated through the measurement model using first-order partial derivatives (sensitivity coefficients) to obtain the combined standard uncertainty of the result [41,42].
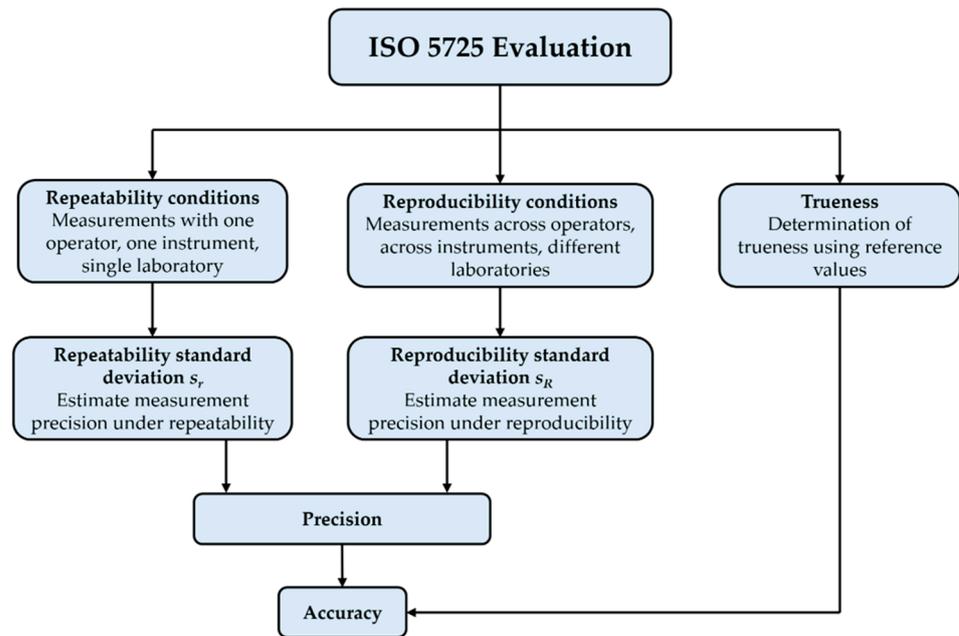
Figure 2. ISO 5725 framework for evaluating measurement accuracy.

The result is typically expanded into an expanded uncertainty by multiplying the combined standard uncertainty by a coverage factor (k), usually $k \approx 2$, to achieve a 95% confidence level. The GUM method assumes that the input quantities are uncorrelated or weakly correlated, that their distributions are normal or approximately normal, and that the model is linear or can be linearized locally [43]. In cases where the model is highly nonlinear or the distributions are non-Gaussian, the Monte Carlo Method (MCM) is recommended as an extension of the GUM approach [44].

The Monte Carlo Method (MCM) is an advanced numerical approach for evaluating measurement uncertainty, particularly suitable for complex or nonlinear measurement models [45–47]. It serves as an extension to the GUM method and is recommended when analytical propagation using partial derivatives becomes impractical or inaccurate [48–50]. Rather than linearizing the measurement model, MCM propagates uncertainty by randomly sampling input quantities from their respective probability distributions and recomputing the output thousands (or millions) of times. This process results in an empirical distribution of the measurand, from which the standard uncertainty and coverage interval can be directly derived.

A key advantage of MCM is its flexibility: it allows for the use of non-Gaussian and asymmetric distributions and nonlinear functions without simplifying assumptions. It produces a coverage interval that reflects the actual shape of the output distribution, including asymmetry and skewness, which is particularly useful when linear approximations would introduce significant errors [51,52].

MCM requires defining input distributions (e.g., normal, uniform, and triangular) based on available data or expert knowledge. The output is typically summarized by the shortest coverage interval at a specified confidence level (e.g., 95%) [53,54]. MCM is increasingly used in metrology institutes, research environments, and high-precision industrial applications, particularly where rigorous uncertainty evaluation is critical.

The Bayesian approach incorporates all the features of the Monte Carlo Simulation (MCS) method but also enables the integration of prior knowledge through informative, weakly informative, or non-informative prior distributions [55,56]. This makes it especially suitable for situations with limited data or when subjective information plays a role. The Bayesian framework typically uses Markov Chain Monte Carlo (MCMC) techniques to

obtain the posterior distribution of the measurand [57–59]. As such, it is particularly well-suited for complex models and scenarios where uncertainty must be fully described through a probabilistic framework.

The Bayesian method for evaluating measurement uncertainty is fundamentally based on Bayes' theorem [60]. In this context, the measurand or parameter of interest is treated as a random variable that follows a certain prior distribution, reflecting existing knowledge or assumptions about its likely values. Measurement results are then used to update this prior distribution, resulting in a posterior distribution that represents the revised state of knowledge after considering the observed data. Importantly, the measurement results and the parameter do not need to originate from the same statistical distribution, allowing for flexible modeling of real-world measurement processes.

The Bayesian procedure involves several conceptual steps: defining the prior distribution based on available knowledge or uncertainty assumptions; specifying a statistical model that relates the observed data to the parameter of interest; and calculating the posterior distribution, which combines the prior information with the evidence provided by the data [61–63]. From this posterior distribution, estimates of the measurand, standard uncertainty, and coverage intervals can be derived. Furthermore, the Bayesian approach allows for the iterative updating of uncertainty estimates as new data become available, making it a powerful tool for adaptive and data-informed measurement uncertainty evaluation, in line with modern metrological practices [64]. A comparison of GUM, Monte Carlo, and Bayesian approaches to uncertainty evaluation is presented in Figure 3.
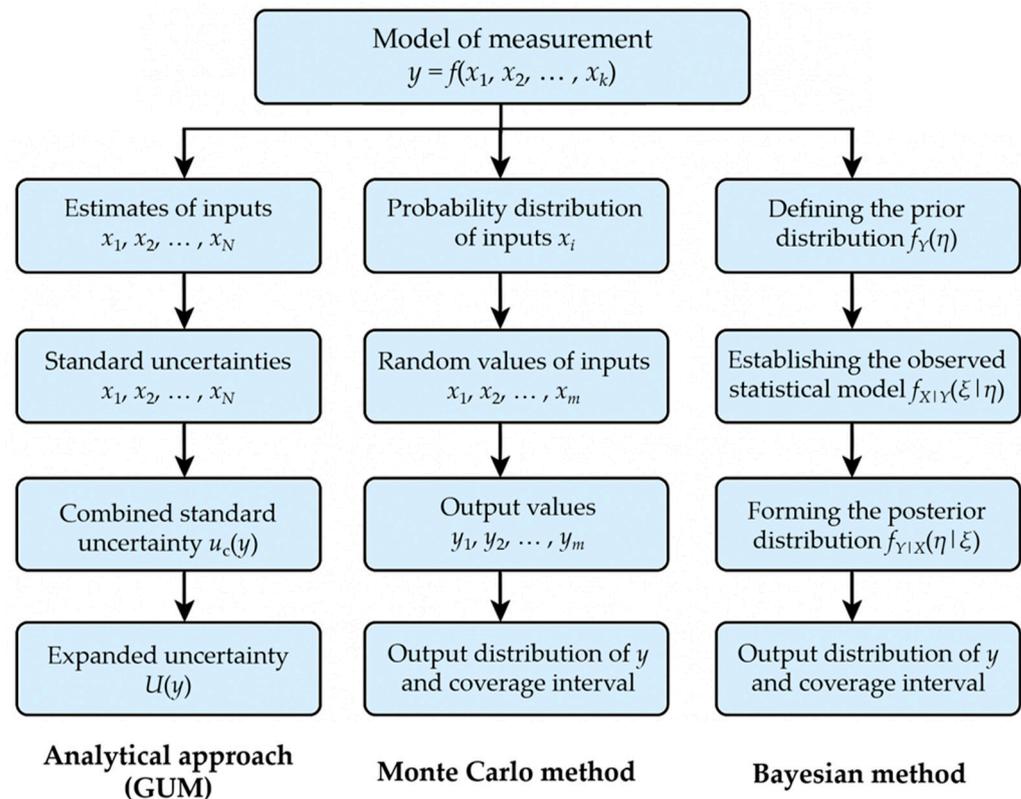
Figure 3. A comparison of the GUM, Monte Carlo, and Bayesian approaches to measurement uncertainty evaluation.

The models and criteria discussed in this paper for the evaluation of measurement uncertainty are broadly applicable and can be adapted to different laboratory conditions and requirements.

ISO 21748:2017 serves as a bridge between ISO 5725 and GUM, offering guidance on how precision and trueness data obtained through structured experiments can be used for uncertainty estimation when a complete uncertainty model is unavailable. This standard is especially valuable for routine laboratories operating under ISO/IEC 17025 that need a practical and traceable approach to uncertainty.

An overview of the methods for assessing the quality of measurement systems and results is illustrated in Figure 4.
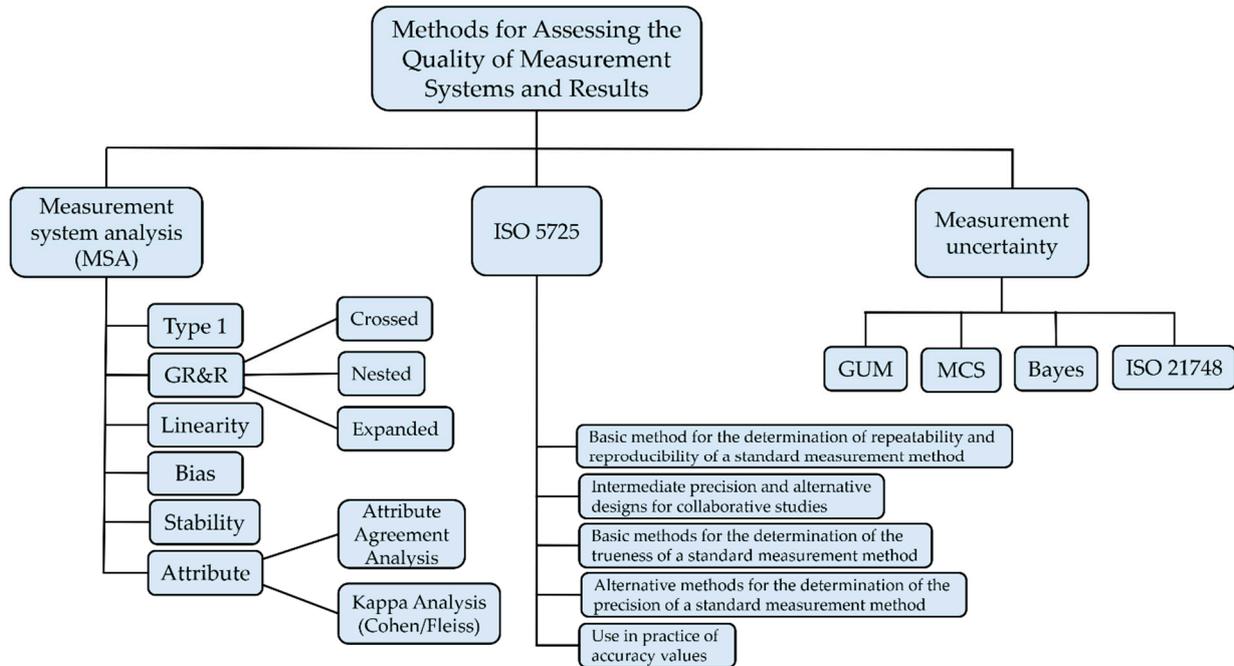

Figure 4. Methods for assessing the quality of measurement systems and results.

3. Methods in Practice

The choice of method for evaluating measurement quality in practice depends on multiple factors, including the purpose of measurement; complexity of the measurement model; data availability; regulatory requirements; and the required level of accuracy, precision and traceability.

3.1. Example of Gage R&R Study Using $\overline{X}$ and R Method

This example demonstrates the application of the Gage R&R ($\overline{X}$ and R) method to evaluate the repeatability and reproducibility of a measurement system. Ten shafts with nominal dimensions of (10.50 ± 0.1) mm were selected to represent the expected range of process variation. Three operators (Operator A, Operator B, and Operator C) measured each of the ten parts three times in random order. The measurements were performed using an outside micrometer. The results are shown in Supplementary Materials. A summary of the Gage R&R results is presented in Table 4.

To evaluate the adequacy of the measurement system, the results are interpreted using three commonly applied criteria given in Table 5.

Table 4. Summary of Gage R&R results.

Source of Variation	%Contribution	%Study Var	%Tolerance
Total Gage R&R	0.32	5.68	12.90
Repeatability	0.26	5.10	11.58
Reproducibility	0.06	2.50	5.69
Part-To-Part	99.68	99.84	226.90
Total Variation	100.00	100.00	227.27

Table 5. Applied criteria.

%Contribution Refers to how much variation is caused by the measurement system itself, in comparison to total variation.	%Study Variation Reflects the proportion of variation due to the measurement system relative to overall process variation.	%Tolerance Compares the variability of the measurement system to the full specification tolerance of the part or product.
≤1% The measurement system is considered acceptable.	<10% Acceptable.	
1–9% May be acceptable, depending on the application.	10–30% Marginally acceptable.	
>9% Unacceptable.	>30% Unacceptable.	

According to %Contribution, the measurement system contributes 0.32% to total variation, which is far below the 1% threshold. This means the system is acceptable.

According to %Study Variation, the Gage R&R is 5.68%, which is below 10%. This indicates that the system is acceptable and contributes minimal variation.

According to %Tolerance, the Gage R&R accounts for 12.90% of the tolerance range, which is also well within acceptable limits. Although a %Tolerance value of 12.90% technically falls within the "Marginally acceptable" range (10–30%) according to standard MSA guidelines, values below 20% are generally considered acceptable in most industrial applications. In practice, the interval between 20% and 30% is subject to further economic justification, evaluating whether the cost and complexity of improving the measurement system outweigh the benefits. In this case, the result of 12.90% indicates a stable and usable measurement system without the need for modification.

Overall conclusion

The measurement system is acceptable. It shows excellent precision, contributes very little measurement variation, and is capable of clearly distinguishing part-to-part differences.

3.2. Example of a Gage R&R Study Using Attribute Agreement Analysis

An Attribute agreement analysis was conducted to examine the consistency of decision-making among three appraisers during the visual inspection of seams on car seats in the ramp-up phase of production (the initial stage of increasing production before full-scale serial production). The aim of the analysis was to confirm the alignment of their decisions with internal quality standards and with reference evaluations provided by experts from the quality control department.

A binary decision criterion was applied in the analysis:

Accept: The seam is straight, without loose threads, and properly symmetrical.

Reject: The seam shows curvature, skipped stitches, or other visible defects.

The three appraisers (A, B, and C) examined the same set of seats, with each appraiser evaluating each sample three times in separate sessions.

At the same time, expert decisions were available and used as reference values for comparison and accuracy analysis.

The analysis was conducted on thirty samples, and the results are presented in Supplementary Materials. Within-appraiser agreement is presented in Table 6.

Table 6. Within appraiser agreement rates and confidence intervals.

Appraiser	Number of Samples	Matched Ratings	Agreement Rate	95% CI
A	30	30	100.00%	(90.50, 100.00)
B	30	29	96.67%	(82.78, 99.92)
C	30	30	100.00%	(90.50, 100.00)

All three appraisers demonstrated very high repeatability, with no significant inconsistencies. Agreement rates ranged from 96.67% to 100.00%, with an average of approximately 98.9%. Appraiser agreement with the reference standard is presented in Table 7.

Table 7. Appraiser agreement with the reference standard.

Appraiser	Number of Samples	Matched Ratings	Agreement Rate	95% CI
A	30	29	96.67%	(82.78, 99.92)
B	30	29	96.67%	(82.78, 99.92)
C	30	28	93.33%	(77.93, 99.18)

Each appraiser individually showed very good agreement with the expert reference evaluations. Disagreement in assessments is presented in Table 8.

Table 8. Assessment disagreement by appraiser.

Appraiser	Incorrect Reject/Accept (%)	Incorrect Accept/Reject (%)	Mixed Ratings (%)
A	0.00	12.50	0.00
B	0.00	0.00	3.33
C	0.00	25.00	0.00

Appraiser C had the highest rate of incorrect acceptance of defective samples (accept instead of reject). Appraiser B had the only mixed rating (inconsistency within repeated assessments). Between-appraiser agreement is presented in Table 9.

Table 9. Between-appraiser agreement.

Number of Samples	Matched Ratings	Agreement Rate	95% CI
30	26	86.67%	(69.28; 96.24)

Out of 30 samples, 4 showed disagreements among appraisers. A total of four disagreements were recorded, aligning with between appraiser's agreement rate of 86.7%. It can

be concluded that appraiser repeatability was very high (average ~98.9%). Inter-appraiser reproducibility was good (86.7%). The overall error rate was 3.7%, which is considered low, though not negligible. The main challenge was identifying defective items (especially for appraiser C). Additional training on borderline samples and alignment of rejection criteria is recommended.

Numerous studies in the literature highlight the wide applicability of Measurement System Analysis (MSA) methods across various industries and measurement scenarios. For example, Abduljaleel et al. [65] explored outlier detection within a nested Gage R&R model using random effects, while Al-Qudah [66] provided a comprehensive overview of the AIAG MSA method in the context of quality control. Al-Refaie and Bata [67] examined the integration of Gage R&R with multiple quality measures to evaluate measurement and process capability.

The use of nested Gage R&R designs for analyzing environmental parameters such as wind speed was presented by Aquila et al. [68], demonstrating the method's flexibility beyond traditional manufacturing. Cepova et al. [69] focused on the implementation of Gage R&R methods in measurement science, while Kooshan [70] described a practical case study involving MSA implementation in the automotive sector.

Advanced approaches such as multivariate Gage R&R have also been investigated. Marques et al. [71] applied factor analysis for multivariate assessments, and Peruchi et al. [72] proposed a novel method for handling correlated characteristics in complex production environments. These examples underline the importance of MSA as a tool for assessing measurement system quality in high-volume industrial production and environmental monitoring.

3.3. Example of the Application of ISO 5725

To illustrate the application of ISO 5725 in assessing the quality of measurement results, we refer to a previously published study that analyzed the repeatability and reproducibility of areal surface topography parameters (Sa, Sq, and Sz) measured by an atomic force microscope (AFM) [73]. In that study, ISO 5725-2:2020 was applied to determine the precision of surface texture parameters across two certified reference standards and a steel sample. Repeatability and reproducibility were evaluated using measurement series conducted under controlled conditions, with changes in measurement points and time intervals used to simulate different laboratory environments. A detailed calculation of repeatability and reproducibility values in accordance with ISO 5725-2:2020 is presented in Table 10. In that table, n denotes the number of repeated measurements within each series; the index i identifies the measurement series ($i = 1, 2, \ldots, m$), while the index j refers to the individual measurement within a given series ($j = 1, 2, \ldots, n$).

The results demonstrated that the AFM measurement system provides high repeatability ($r < 1$ nm) and reproducibility ($R < 2$ nm) for parameters Sa and Sq, while parameter Sz showed greater sensitivity to measurement location. This example shows how ISO 5725 can be applied beyond inter-laboratory comparisons, by adapting the methodology for intra-laboratory precision studies, which is especially useful in specialized or unique laboratory environments. In accordance with ISO 21748, the calculated values of repeatability and reproducibility are treated as constituent components of the measurement uncertainty associated with the respective reference standards.

Table 10. Statistical evaluation of repeatability and reproducibility for *Sa*, *Sz*, and *Sq* parameters according to ISO 5725-2:2020 (measurement years 2021 and 2025).

	Sa/nm		*Sz*/nm		*Sq*/nm	
Year 2021	Measurement series		Measurement series		Measurement series	
	1	2	1	2	1	2
$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n} x_{ij}$	50.849	50.157	174.521	175.914	54.646	54.385
$s_i = \sqrt{\frac{\sum_{j=1}^{n}(x_i - \overline{x})^2}{n_i - 1}}$	0.151	0.224	1.395	1.590	0.468	0.202
$\overline{\overline{x}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}$	50.503		175.218		54.515	
$s_r = \sqrt{\frac{\sum_{i=1}^{m}(n_i-1)s_i^2}{\sum_{i=1}^{m} n_i - 1}}$	0.191		1.495		0.360	
$s_d = \sqrt{\frac{\sum_{i=1}^{m}(\overline{x}_i - \overline{\overline{x}})^2}{m-1}}$	0.346		0.697		0.130	
$s_L = \sqrt{s_d^2 - \frac{s_r^2}{n}}$	0.342		0.580		0.091	
$s_R = \sqrt{s_r^2 + s_L^2}$	0.392		1.604		0.372	
$r = 2.8\,s_r$	0.528		4.133		0.995	
$R = 2.8\,s_R$	1.083		4.433		1.027	
Year 2025	Measurement series		Measurement series		Measurement series	
	1	2	1	2	1	2
$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n} x_{ij}$	50.24	51.24	182.50	176.63	51.812	53.71
$s_i = \sqrt{\frac{\sum_{j=1}^{n}(x_i - \overline{x})^2}{n_i - 1}}$	0.16	0.19	2.03	1.63	0.174	0.244
$\overline{\overline{x}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}$	50.74		179.63		52.76	
$s_r = \sqrt{\frac{\sum_{i=1}^{m}(n_i-1)s_i^2}{\sum_{i=1}^{m} n_i - 1}}$	0.177		1.843		0.212	
$s_d = \sqrt{\frac{\sum_{i=1}^{m}(\overline{x}_i - \overline{\overline{x}})^2}{m-1}}$	0.503		2.064		0.947	
$s_L = \sqrt{s_d^2 - \frac{s_r^2}{n}}$	0.122		0.240		0.238	
$s_R = \sqrt{s_r^2 + s_L^2}$	0.215		1.859		0.319	
$r = 2.8\,s_r$	0.489		5.094		0.586	
$R = 2.8\,s_R$	0.593		5.394		0.881	

Although ISO 5725 and ISO 21748 provide a structured and standardized approach for evaluating measurement accuracy and precision, as well as for estimating uncertainty based on repeatability and reproducibility data, their practical application in the contemporary scientific literature remains limited. Given the very small number of available publications, there is a clear lack of recent research that systematically integrates the methodologies of ISO 5725 and ISO 21748. Among the more recent works, the study by Bontempi et al. [74] can be highlighted, where the authors applied ISO 5725 in the context of software validation for radiostereometric analysis. D'Aucelli et al. [75] developed a MATLAB framework to

support MSA evaluations based on ISO 5725. The Singapore Accreditation Council's (SAC) Technical Guide 2—Guidelines on Measurement Uncertainty for Testing Laboratories (2023) provides practical guidance for evaluating measurement uncertainty in accordance with ISO/IEC 17025. The guide explicitly builds upon the methodologies outlined in ISO 5725 for assessing precision (repeatability and reproducibility) and refers to ISO 21748 for using such precision data to estimate measurement uncertainty in the context of routine testing [76]. A publication that provides a detailed description of the main objectives and methods of ISO 5725 precision, trueness, and the implementation of interlaboratory comparisons is the paper titled "*Basics of interlaboratory studies: the trends in the new ISO 5725 standard*" by Max Feinberg [77], which dates back to 1995.

This significant time gap between the foundational work and current technological developments highlights the need for broader adoption, practical case studies, and updated guidance documents and examples tailored to modern laboratory and industrial environments.

3.4. Example of Measurement Uncertainty Evaluation Using the GUM Method

The thermometer is calibrated by comparing $n = 11$ temperature readings from the thermometer with corresponding known reference temperatures in the range from 21 °C to 27 °C [10]. The linear calibration curve is given by

$$b(t) = y_1 + y_2(t - t_0), \tag{1}$$

where

- $b(t)$ is the predicted value of the thermometer correction at temperature t, expressed in °C;
- $y_1 = -0.1712$ °C is the intercept of the calibration curve, with an associated standard uncertainty of $s(y_1) = 0.0029$ °C;
- $y_2 = 0.00218$ °C is the slope of the calibration curve, with an associated standard uncertainty of $s(y_2) = 0.00067$ °C;
- $r(y_1, y_2) = -0.930$ is the estimated correlation coefficient between y_1 and y_2;
- $s = 0.0035$ °C is the estimated standard deviation of the fit residuals;
- t is the temperature at which the correction is to be predicted, expressed in °C;
- t_0 is the chosen reference temperature, an exactly known fixed value, expressed in °C.

The given exact reference temperature is $t_0 = 20$ °C. Thus, the calibration curve can be written as

$$b(t) = -0.1712 \text{ °C} + 0.00218 \text{ °C} \, (t - 20 \text{ °C}) \tag{2}$$

The combined standard uncertainty is obtained by taking the partial derivatives of each term and is given by

$$u_c^2[b(t)] = u^2(y_1) + (t - t_0)^2 u^2(y_2) + 2(t - t_0)u(y_1)u(y_2)r(y_1, y_2) \tag{3}$$

The correction of the thermometer reading and its uncertainty at $t = 30$ °C, which is outside the temperature range in which the thermometer was actually calibrated, is $b(30 \text{ °C}) = -0.1494$ °C. Combined standard uncertainty is $U_c[b(30 \text{ °C})] = 0.0041$ °C. Expanded measurement uncertainty is calculated as

$$U_c[b(30 \text{ °C})] = k \, u_c[b(30 \text{ °C})] = 2 \times 0.0041 \text{ °C} = 0.0082 \text{ °C} \tag{4}$$
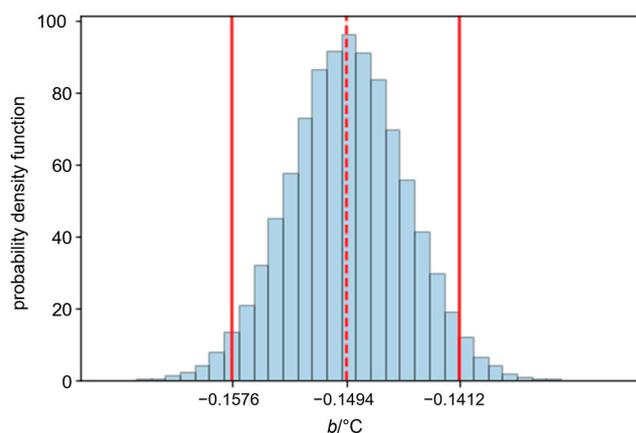
3.5. Example of Measurement Uncertainty Evaluation Using the Monte Carlo Simulation Method

The same example, using the same input data, was evaluated using the Monte Carlo simulation method in accordance with the GUM Supplement 1 guidelines (JCGM 101) [11]. The mathematical model describing the relationship between the input and output quantities is given in equation (1). A total of $M = 100,000$ simulations were carried out to model the propagation of uncertainty through the calibration function. The input quantities and their associated probability distributions used in the Monte Carlo simulation are summarized in Table 11.

Table 11. Input quantities and probability density function.

Quantity	Symbol	Distribution	Value
Temperature	t	Fixed value	30 °C
Reference temperature	t_0	Constant	24.0085 °C
Intercept (calibration curve)	y_1	Normal	$(M, -0.1625, 0.0011)$
Slope (calibration curve)	y_2	Normal	$(M, 0.00218, 0.00067)$

Based on the mathematical model and the specified input quantities, the Monte Carlo simulation yields the probability distribution of the correction corresponding to the selected temperature. In Figure 5, the arithmetic mean and the boundaries of the 95% coverage interval are indicated. The expanded measurement uncertainty is $U = 0.00816$ °C.

Figure 5. Monte Carlo simulations method: probability density function.

The Monte Carlo simulation method confirmed the result obtained using the GUM analytical approach, demonstrating consistency between the two methods. Moreover, the MC approach provides a full probability distribution of the result, which is particularly useful when input distributions are non-normal or the model is nonlinear.

3.6. Example of Measurement Uncertainty Evaluation Using the Bayesian Approach

A significant foundation for the approach presented in this section stems from the doctoral research of one of the authors, which investigates the evaluation of measurement uncertainty in AFM-based surface metrology [78]. The application of the Bayesian method is demonstrated through the evaluation of the measurement uncertainty of the step height of an AFM reference standard. The calibration certificate for the standard, with a nominal value of 100 nm, states that the step height is (97.6 ± 1.4) nm, with a coverage probability $p = 95\%$. In the Bayesian approach, the certified value of the step height is incorporated as

prior information in the uncertainty evaluation process. In the paper [79], a mathematical model of the step height is given as follows:

$$h = h_x + r \cdot t \cdot l + d \cdot \alpha \cdot \Delta\vartheta + \delta r + \delta R + \delta probe \tag{5}$$

The variables used in the model and their physical meanings are summarized as follows: h_x represents the measured value (indication); r, the scan rate; t, the scanning time; l, the scan length; d, the nominal step height; α, the temperature expansion coefficient; $\Delta\vartheta$, the temperature difference; δr, the repeatability component; δR, the reproducibility component; and $\delta probe$, the probe-related uncertainty. When applying the Bayesian method, the mathematical model of the measured variable (5) needs to be transformed into an observational model. The observational model is given by expression (6).

$$h_x = h - (r \cdot t \cdot l + d \cdot \alpha \cdot \Delta\vartheta + \delta r + \delta R + \delta probe), \tag{6}$$

where $D \equiv h_x = \phi(Y, \Theta)$ represents the indication size, $Y \equiv h$ represents the output variable, $\Theta \equiv (v, t, l, h, \alpha, \Delta T, \delta r, \delta R, \delta probe)$ represents other input variables.

The indication values represent the results of 20 repeated measurements with an expected value of D and a standard uncertainty S. Prior distributions of input variables are created based on the information from [80]. In the implementation of the Bayesian method for the output variable, a less informative prior distribution was used (the value of step height from the calibration certificate). The output variable obtained using the Bayesian method is illustrated by Figure 6: the cumulative distribution function and the probability density function.
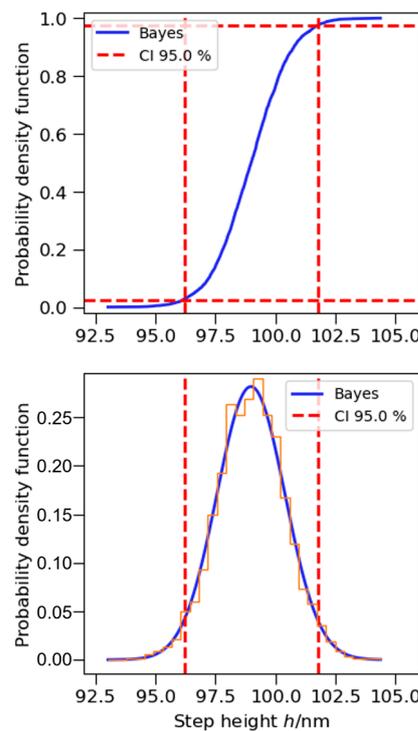

Figure 6. Bayesian method: step height h [78].

The output variable h is within the interval ($Y_{0.025}$ = 96.5 nm, $Y_{0.975}$ = 100.2 nm) with p = 95%. The comparison of results obtained by the Monte Carlo simulation method calculated in [80] and the Bayesian method is provided in Table 12.

Table 12. Results for symmetric coverage interval (95%).

Method	h	u(h)	Coverage Interval (95%)
	nm	nm	nm
MCS	98.9	1.45	[96.0, 101.8]
Bayes	98.3	1.41	[96.1, 101.7]

Reasonable agreement of the results is shown, with differences less than 5%. The smaller uncertainty obtained using the Bayesian method compared to the Monte Carlo simulation method can be attributed to the use of prior information about the output quantity.

In addition to the calibration examples presented in this paper, a variety of well-documented case studies illustrating good practice in measurement uncertainty evaluation can be found in the publication "Good Practice in Evaluating Measurement Uncertainty—Compendium of Examples" [81], edited by Van der Veen and Cox and published by EURAMET. The compendium includes examples across different measurement domains and illustrates the application of the GUM method, the Monte Carlo method, and Bayesian approaches.

For example, the GUM method is applied in "Two-point and multipoint calibration" (Cox et al. [82]), which addresses pH measurement and emphasizes the importance of input quantity correlations. Another GUM-based example is "Evaluation of measurement uncertainty in average areal rainfall" (Ribeiro et al. [83]), which compares multiple estimation methods using GUM and GUM Supplement 1.

The Monte Carlo method is used in "Evaluation of measurement uncertainty in SBI—Single Burning Item reaction to fire test" (Martins et al. [84]), where the complexity and nonlinearity of the model justify the use of simulation-based propagation. A more extensive application is shown in "Greenhouse gas emission inventories" (Cox et al. [85]), which combines MCS and Bayesian methods in environmental modeling.

The Bayesian approach is demonstrated in the example "Bayesian approach applied to the mass calibration example in JCGM 101:2008" (Demeyer et al. [86]), where prior knowledge is incorporated into the evaluation of a 100 g mass standard. A more complex application is found in "Uncertainty evaluation of nanoparticle size by AFM" (Caebergs et al. [87]), which uses a hierarchical Bayesian model with an optimized Design of Experiment to efficiently support nanometrology applications.

These examples serve as valuable references when designing, validating, or improving uncertainty evaluation strategies in laboratory practice.

4. Discussion: Strengths, Limitations, and Method Selection

4.1. Applicability and Scope of Measurement System Analysis (MSA)

The presented approaches can be compared, but they are not interchangeable; each has its own context of application. It is often desirable for them to complement each other within larger measurement management systems. MSA (Measurement System Analysis) is most appropriately used in industrial environments, for both variable and attribute data.

The methods for variable characteristics within the MSA framework are applied:

- To assess the quality of measurement systems in order to understand how much of the variation in results is caused by the measurement system compared to the actual product variation.
- To analyze components within the measurement system (gauge, operator, and interactions).
- When introducing new measurement systems, to verify their suitability before use.
- In the context of quality control when measuring or verifying product compliance with requirements.

The methods for attribute characteristics within the MSA framework are applied:

- In situations where discrete decisions are made (e.g., during inspection or visual examination).
- Multiple operators evaluate the same set of parts (e.g., pass/fail) multiple times, and the results are compared with each other and with a known "gold standard". The following aspects are analyzed: accuracy relative to the reference (how often the operator agrees with the correct answer), internal consistency of the operator (repeatability), and consistency among operators (reproducibility).
- To assess inter-rater agreement, especially in situations where multiple rating levels are possible (e.g., low/medium/high defect severity). The kappa coefficient indicates the degree of agreement beyond chance.

4.2. Advantages and Limitations of the MSA

Advantages of the MSA approach:

- It is ideal for quick, operational verification of measurement capability within industrial production.
- It clearly links the measurement system to process control and decision-making—for example, whether a result can be used to accept or reject a part.
- It provides clear thresholds and acceptability criteria.
- MSA is an integrated part of quality management systems and tools such as SPC (Statistical Process Control), PPAP (Production Part Approval Process), and APQP (Advanced Product Quality Planning).
- The focus is on the measurement system as a whole, not just the measurement result.

Limitations of the MSA approach:

- It is tailored to specific industrial processes and is less applicable in scientific or research environments.
- It requires multiple operators to repeat measurements on the same sample, which is often not feasible in practice.
- It does not provide a systematic evaluation of measurement uncertainty.
- It is limited to production environments and intra-laboratory assessments.

4.3. Applicability and Scope of the ISO 5725 Series of Standards

The ISO 5725 series of standards is used:

- For designing and conducting interlaboratory studies aimed at determining the repeatability (within-laboratory) and reproducibility (between-laboratories) of measurement results.
- For evaluating the trueness of measurement methods based on experimental data obtained from multiple laboratories.
- For internal quality control of measurement methods within a single laboratory, especially when the same method is applied continuously over time.
- For assessing the accuracy (i.e., trueness and precision) of a method in research settings where the number of available samples is small, measurement conditions are not standardized, or the method is applied in a non-standard context.

Advantages of the Approach Based on the ISO 5725 Series of Standards:

- ISO 5725 covers a wide range of measurement situations, allowing for the analysis of results obtained under different conditions, from various laboratories and methods.
- The approach is internationally standardized, meaning that the results and procedures are globally recognized and accepted, especially in the scientific community, in industry, and by regulatory bodies.

- The standards distinguish between different sources of variability, including repeatability, intermediate precision, and reproducibility, enabling a detailed analysis of method performance under varying conditions.
- The approach can be adapted to different levels of complexity, from simple laboratory tests to full interlaboratory studies, depending on the user's needs.

Limitations of the Approach Based on the ISO 5725 Series of Standards:

- It requires a solid understanding of statistics (e.g., ANOVA, variance analysis, and outlier tests), which can be a barrier for use in small laboratories or in industrial settings without data specialists.
- It does not provide numerical values of measurement uncertainty directly but focuses on estimates of precision and trueness as components of accuracy. As a result, results often need to be further converted into uncertainty estimates for accreditation purposes.
- It is not designed for rapid evaluation of measurement systems on production lines, where quick detection of operator or instrument deviations is crucial.
- The standards were originally developed for analytical and laboratory methods (e.g., chemical and microbiological), making them more difficult to apply directly to methods involving complex image processing (e.g., AFM, 3D optical measurements, and Computed Tomography).
- They are not designed to systematically analyze and control all digital and software-related factors that may influence the measurement result.

4.4. Applicability and Scope of Measurement Uncertainty Evaluation

Measurement uncertainty evaluation is a key component in all fields where decisions are made based on measurement results. The most commonly used approaches, GUM, the Monte Carlo method, and the Bayesian approach, are widely applied in testing and calibration laboratories, especially those operating under ISO/IEC 17025, which requires a clearly documented and justified uncertainty evaluation for all test and calibration results.

In the field of medical laboratories, ISO 15189 [88] requires the evaluation of measurement uncertainty for quantitative analyses, such as determining the concentration of glucose, hormones, or electrolytes. Given the specific nature of medical measurements, the GUM-based approach is often adapted to the clinical context to enable reliable interpretation of results critical for diagnosis and treatment.

Inspection bodies operating in accordance with ISO/IEC 17020 [89] must also evaluate measurement uncertainty when measurement results serve as the basis for inspection decisions—for example, in the assessment of the safety of elevators, pressure vessels, or industrial pipelines. ISO 10012 [90], which relates to measurement management systems in industrial and manufacturing environments, recommends the management of measurement uncertainty to ensure measurement traceability and product quality. Similarly, the ISO 14253 [91] series of standards, which address product geometric specifications (GPS), emphasize the need to incorporate measurement uncertainty when deciding on the conformity of measured features with technical requirements, an aspect particularly important in dimensional metrology.

Although ISO 9001 [92] does not explicitly mandate the evaluation of measurement uncertainty, it is implied in the requirements for ensuring the quality of measurement systems, especially for manufacturers operating in high-precision industries. Demonstrating that the measurement system is capable of meeting the required accuracy necessarily involves understanding and managing measurement uncertainty.

The evaluation of measurement uncertainty is also mandatory according to the accreditation requirements of international bodies such as International Laboratory Accreditation

Cooperation (ILAC), German Accreditation Body (DAkkS), Croatian Accreditation Agency (HAA), United Kingdom Accreditation Service (UKAS), and other national accreditation agencies. For accredited testing and calibration laboratories, reporting measurement uncertainty is considered an essential component of reliable and traceable measurement, particularly in calibration certificates, where omitting uncertainty would render the result incomplete.

In addition, measurement uncertainty plays an important role in scientific research, serving as a foundation for the credible interpretation of experimental results. Without quantifying uncertainty, scientific claims regarding the significance of differences, repeatability of results, or conformity with theoretical models remain questionable. In research involving measurements, ranging from physics and chemistry to engineering and biomedicine, uncertainty is crucial for transparency, comparability, and reproducibility of the data obtained. In an increasing number of scientific journals, including uncertainty evaluation is regarded as good scientific practice or even a formal requirement for publication.

4.5. Advantages and Limitations of GUM, Monte Carlo, and Bayesian Methods

4.5.1. Advantages of the GUM, Monte Carlo, and Bayesian Approaches

In contrast to traditional Measurement Systems Analysis (MSA) methods, such as Gage R&R studies or ISO 5725-based approaches, which primarily rely on statistical analysis of experimental data to assess repeatability, reproducibility, bias, and precision, modern approaches to evaluating measurement uncertainty—namely GUM, Monte Carlo Simulation (MCS), and Bayesian methods—offer several important advantages.

First, the GUM (Guide to the Expression of Uncertainty in Measurement) approach enables analytical modeling of the measurement process through a mathematical measurement equation, incorporating all known sources of uncertainty, whether derived from experimental measurements, calibration certificates, equipment specifications, or expert judgment. This approach provides a structured and traceable evaluation of uncertainty, even when extensive experimental studies are not feasible.

Monte Carlo Simulation (MCS) extends the GUM framework for nonlinear measurement models or situations where an analytical solution is not possible. Instead of uncertainty propagation via differentiation, MCS uses numerical simulation of the result distribution, enabling a more realistic depiction of measurement uncertainty, especially in complex or nonlinear systems. MCS is particularly valuable in advanced engineering and scientific applications where traditional statistical models may not be suitable.

The Bayesian approach offers additional flexibility by allowing for the combination of prior knowledge (e.g., from previous calibrations, historical data, or expert judgment) with new experimental data. The outcome is not only an estimate of uncertainty but also an updated probability distribution of the measurement. This enables dynamic improvement of the accuracy and reliability of measurement results, which is especially useful in situations with a limited number of samples or in sequential measurements.

These three approaches, GUM, MCS, and Bayesian, represent a fundamental shift from descriptive statistics to interpretative and predictive modeling of measurement uncertainty, making them incomparably more powerful tools for today's demanding measurement tasks.

4.5.2. Limitations of the GUM, Monte Carlo, and Bayesian Approaches

Although modern approaches to measurement uncertainty evaluation—such as the GUM methodology, Monte Carlo simulation, and the Bayesian approach—offer deeper insights and greater flexibility in modeling measurement processes, they are not without

limitations, especially when compared to classical statistical methods such as MSA and ISO 5725.

One of the main limitations of the GUM approach is the requirement for a detailed understanding of the measurement model, including all relevant quantities, their interrelationships, and associated sources of uncertainty. Unlike ISO 5725, which relies on experimental repetition and does not require explicit modeling, GUM demands a mathematical formulation of the measurement process, which can be challenging in complex or poorly understood systems.

The Monte Carlo method, although powerful for nonlinear problems, requires computational resources and statistical expertise, which may be a barrier in laboratories lacking advanced software tools or trained personnel. Furthermore, the quality of the results depends on the number of simulations and the validity of the input distributions, meaning that incorrect assumptions can lead to unreliable outcomes.

The Bayesian approach introduces additional complexity, as it requires the definition of prior distributions, which may introduce subjectivity if prior knowledge is not well-founded. Furthermore, its practical application is not yet widely standardized and often involves advanced computational methods (e.g., MCMC), which are not typically part of routine laboratory practice.

In summary, although GUM, MCS, and Bayesian methods offer greater theoretical precision and flexibility, their application is more demanding. They depend on the availability of data, expert knowledge, and computational resources and may not always be practical in industrial or operational contexts where speed, simplicity, and interpretability are more important than complete analytical accuracy.

4.6. Limitations and Sources of Uncertainty

This review paper includes experimental data and numerical implementations of methods for evaluating measurement quality, conducted in both industrial and laboratory environments. Methods such as Gage R&R, Attribute Agreement Analysis, GUM, Monte Carlo, and the Bayesian approach were applied to real-world data, enabling practical comparison of the approaches. Although this study is based on a large number of selected relevant sources, it is still possible that some pertinent studies or perspectives were omitted from the analysis. Due to the lack of comparable quantitative studies in the literature, the analysis is based on a structured comparison of methodological approaches and their characteristics.

While this paper is grounded in actual measurement data, the scope of experimental examples is limited. This study focuses on specific cases within certain metrological domains. The analyses and results do not cover all possible industrial and laboratory scenarios; therefore, further validation in other contexts would be required before generalizing the conclusions. The comparison of methods and standards (MSA, ISO 5725, GUM, etc.) is based on current standards and expert sources. However, differences in terminology and interpretation of key concepts (e.g., "accuracy", "trueness", and "repeatability") still pose challenges for full alignment among the approaches.

Despite these limitations, this paper contributes to a better understanding of available methods and standards for assessing measurement quality and may serve as a foundation for further validation and application in various metrological contexts.

5. Conclusions

This review has presented key methodologies for evaluating the quality of measurement systems and results, including Measurement System Analysis (MSA), ISO 5725, and GUM-based uncertainty evaluation. Each approach has its strengths and is suitable for dif-

ferent measurement contexts from industrial production control to high-precision scientific measurements. Through comparative analysis and practical examples, it is demonstrated that a clear understanding of concepts such as repeatability, reproducibility, trueness, accuracy, bias, and measurement uncertainty is essential for correctly interpreting results and applying methods within each specific metrological framework. Since these terms are defined and evaluated differently across various standards (e.g., MSA, ISO 5725, and GUM), this paper examines these concepts in the context of each framework, highlighting their methodological distinctions.

The review has also highlighted inconsistencies in definitions and the use of key terms across different standards and methodological frameworks. For instance, the concept of accuracy is interpreted differently in MSA and ISO 5725, while the term trueness is absent from GUM, despite its central role in ISO 5725. Particularly notable is the inconsistent treatment of bias: in ISO 5725, it represents a measurable component of systematic error used to assess trueness, while in GUM, known biases are corrected and only the uncertainty of the correction is propagated. In contrast, MSA sometimes ambiguously equates bias with accuracy, which can lead to conceptual confusion. Such discrepancies can lead to misinterpretation of results or the misuse of methods if not carefully considered. The ongoing revision of the International Vocabulary of Metrology (VIM), currently in the form of VIM4 Committee Draft (2021), confirms the need for clearer and harmonized definitions. This revision seeks to align terminology across standards such as ISO 5725 and GUM and to improve the consistency of metrological communication and reporting.

In this context, careful interpretation of measurement results is not only a matter of technical correctness but also of terminological precision. Continued efforts toward standard harmonization, education on terminology, and method selection criteria are necessary to ensure reliable and traceable quality assessment particularly in light of future trends involving big data, AI, and intelligent measurement systems.

6. Future Directions and Recommendations

Given the increasing complexity of measurement systems, the growing volume of available data, and the rapid advancement of digital technologies, it is expected that methods for evaluating measurement quality will undergo significant evolution in the near future.

In the context of Industry 4.0 and 5.0, and the ongoing digital transformation of manufacturing, as illustrated in research [93–95], there is a growing need to develop smart measurement systems fully integrated with autonomous robotic platforms capable of making decisions in real time.

In addition to new hardware development, it is essential to standardize methods for evaluating measurement uncertainty in automated and flexible manufacturing environments and to implement advanced software tools for automatic inclusion of uncertainty in reported results. Furthermore, there is increasing interest in applying Artificial Intelligence (AI) and Machine Learning (ML) algorithms in the analysis of measurement system quality. Automated pattern recognition, identification of error sources, and intelligent optimization of measurement procedures offer substantial potential for improving efficiency and reliability.

Uncertainty quantification for supervised machine learning is a relatively new scientific field focused on integrating metrological principles into machine learning. Uncertainty quantification in supervised learning is becoming increasingly important in the context of large-scale data and the deployment of models in sensitive domains such as medicine [96–98]. In decision-making scenarios, such as diagnostics, risk assessment, or personalized therapy, it is crucial to know how "confident" a model is in its prediction [99].

Given the growing volume of heterogeneous data (big data) and the widespread adoption of AI systems, uncertainty quantification is expected to become a key component of future research in machine learning, especially in areas where incorrect decisions carry significant risk.

It is therefore recommended that future efforts focus on

- Continued education of engineers and measurement system users on the importance of expressing and interpreting measurement uncertainty;
- Promotion of the integration of advanced numerical and probabilistic methods into everyday practice;
- Exploration of the potential of AI and ML in metrology;
- Standardization of uncertainty evaluation procedures for automated production environments;
- Development of modular tools and systems for quality assessment, enabling the flexible application of different methods (GUM, MCS, and Bayesian) according to user needs.

The combination of harmonized terminology, robust methods, and advanced technologies will ensure sustainable and highly reliable quality evaluation of measurements in increasingly demanding industrial and scientific settings.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app15179393/s1.

Author Contributions: Conceptualization, B.R., A.R. and V.A.; methodology, A.R., B.R. and V.A.; software, Z.T. and A.R.; resources, B.Š. and B.R.; writing—original draft preparation, A.R., B.R. and B.Š.; resources, V.A. and. A.R.; validation, Z.T. and B.Š.; writing—review and editing, V.A., B.Š. and Z.T.; formal analysis, B.R., B.Š. and Z.T.; investigation, A.R. and B.R.; visualization, Z.T. and V.A.; supervision, B.Š. and B.R. All authors have read and agreed to the published version of the manuscript.

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

References

1. Pendrill, L.R. Using Measurement Uncertainty in Decision-Making and Conformity Assessment. *Metrologia* **2014**, *51*, S206. [CrossRef]
2. BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPA; OIML. *Evaluation of Measurement Data—The Role of Measurement Uncertainty in Conformity Assessment*; JCGM: Sèvres, France, 2012.
3. AIAG. *Measurement Systems Analysis: Reference Manual*, 4th ed.; Automotive Industry Action Group: Southfield, MI, USA, 2010.
4. *ISO 5725-1:2023*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions. International Organization for Standardization: Geneva, Switzerland, 2023.
5. *ISO 5725-2:2019*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method. International Organization for Standardization: Geneva, Switzerland, 2019.
6. *ISO 5725-3:2023*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 3: Intermediate Precision and Alternative Designs for Collaborative Studies. International Organization for Standardization: Geneva, Switzerland, 2023.
7. *ISO 5725-4:2020*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 4: Basic Methods for the Determination of the Trueness of a Standard Measurement Method. International Organization for Standardization: Geneva, Switzerland, 2020.
8. *ISO/DIS 5725-5*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 5: Alternative Methods for the Determination of the Precision of a Standard Measurement Method. International Organization for Standardization: Geneva, Switzerland, 1998.
9. *ISO 5725-6:1994*; Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 6: Use in Practice of Accuracy Values. International Organization for Standardization: Geneva, Switzerland, 1994.

10. *JCGM 100:2008*; Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. International Organization for Standardization: Geneva, Switzerland, 2008.
11. *JCGM 101:2008*; Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method. International Organization for Standardization: Geneva, Switzerland, 2008.
12. Cox, M.; Shirono, K. Informative Bayesian Type A Uncertainty Evaluation, Especially Applicable to a Small Number of Observations. *Metrologia* **2017**, *54*, 642–652. [CrossRef]
13. Van Der Veen, A.M.H. Bayesian Methods for Type A Evaluation of Standard Uncertainty. *Metrologia* **2018**, *55*, 670–684. [CrossRef]
14. Lira, I.; Grientschnig, D. Bayesian Assessment of Uncertainty in Metrology: A Tutorial. *Metrologia* **2010**, *47*, R1–R14. [CrossRef]
15. *ISO 21748:2017*; Guidance for the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Evaluation. International Organization for Standardization: Geneva, Switzerland, 2017.
16. Bich, W.; Cox, M.G.; Dybkaer, R.; Elster, C.; Estler, W.T.; Hibbert, B.; Imai, H.; Kool, W.; Michotte, C.; Nielsen, L.; et al. Revision of the '*Guide to the Expression of Uncertainty in Measurement*'. *Metrologia* **2012**, *49*, 702–705. [CrossRef]
17. Arendacká, B.; Täubner, A.; Eichstädt, S.; Bruns, T.; Elster, C. Linear Mixed Models: Gum and Beyond. *Meas. Sci. Rev.* **2014**, *14*, 52–61. [CrossRef]
18. Runje, B.; Novak, A.H.; Razumić, A.; Piljek, P.; Štrbac, B.; Orosnjak, M. Evaluation of Consumer and Producer Risk in Conformity Assessment Decisions. In Proceedings of the 30th Daaam International Symposium on Inteligent Manufacturing and Automation, Zadar, Croatia, 23–26 October 2019; Danube Adria Association for Automation and Manufacturing, DAAAM: Vienna, Austria, 2019.
19. Cox, M.G.; Harris, P.M. Software Support for Metrology Best Practice Guide No. 6. Uncertainty Evaluation. Available online: https://eprintspublications.npl.co.uk/4655/ (accessed on 24 July 2025).
20. Kacker, R.N. Bayesian Alternative to the ISO-GUM's Use of the Welch–Satterthwaite Formula. *Metrologia* **2005**, *43*, 1. [CrossRef]
21. Aggogeri, F.; Barbato, G.; Barini, E.M.; Genta, G.; Levi, R. Measurement Uncertainty Assessment of Coordinate Measuring Machines by Simulation and Planned Experimentation. *CIRP J. Manuf. Sci. Technol.* **2011**, *4*, 51–56. [CrossRef]
22. Božić, D.; Runje, B.; Razumić, A. Risk Assessment for Linear Regression Models in Metrology. *Appl. Sci.* **2024**, *14*, 2605. [CrossRef]
23. Forbes, A. Traceable Measurements Using Sensor Networks. *Trans. Mach. Learn. Data Min.* **2015**, *8*, 77–100.
24. Bell, S.A. *A Beginner's Guide to Uncertainty of Measurement*; National Physical Laboratory: Teddington, UK, 2001.
25. *ISO-ISO/IEC 17025*; Testing and Calibration Laboratories. International Organization for Standardization: Geneva, Switzerland, 2017.
26. Forbes, A.B. Measurement Uncertainty and Optimized Conformance Assessment. *Measurement* **2006**, *39*, 808–814. [CrossRef]
27. Allard, A.; Fischer, N.; Smith, I.; Harris, P.; Pendrill, L. Risk Calculations for Conformity Assessment in Practice. In Proceedings of the 19th International Congress of Metrology (CIM2019), Paris, France, 24–26 September 2019; EDP Sciences: Les Ulis, France, 2019; p. 16001.
28. Ellison, S.L.R.; Williams, A. *Quantifying Uncertainty in Analytical Measurement*, 3rd ed.; Eurachem/CITAC: Teddington, UK, 2012.
29. Burdick, R.K.; Borror, C.M.; Montgomery, D.C. A Review of Methods for Measurement Systems Capability Analysis. *J. Qual. Technol.* **2003**, *35*, 342–354. [CrossRef]
30. Zanobini, A.; Sereni, B.; Catelani, M.; Ciani, L. Repeatability and Reproducibility Techniques for the Analysis of Measurement Systems. *Measurement* **2016**, *86*, 125–132. [CrossRef]
31. Simion, C. Assessing Bias and Linearity of a Measurement System. *Appl. Mech. Mater.* **2015**, *760*, 621–626. [CrossRef]
32. Pai, F.-Y.; Yeh, T.-M.; Hung, Y.-H. Analysis on Accuracy of Bias, Linearity and Stability of Measurement System in Ball Screw Processes by Simulation. *Sustainability* **2015**, *7*, 15464–15486. [CrossRef]
33. Abdelgadir, M.; Gerling, C.; Dobson, J. Variable Data Measurement Systems Analysis: Advances in Gage Bias and Linearity Referencing and Acceptability. *Int. J. Metrol. Qual. Eng.* **2020**, *11*, 16. [CrossRef]
34. Sahay, C.; Ghosh, S.; Haja Mohideen, S.M. Effects of Within Part Variation on Measurement System Analysis. In Proceedings of the ASME International Mechanical Engineering Congress and Exposition, Houston, TX, USA, 9–15 November 2012; American Society of Mechanical Engineers: New York, NY, USA, 2012; Volume 45196, pp. 1827–1836.
35. Runje, B.; Novak, A.H.; Razumić, A. Measurement System Analysis in Production Process. In Proceedings of the XVII International Scientific Conference on Industrial Systems (IS'17), Novi Sad, Serbia, 4–6 October 2017; pp. 274–277.
36. Wang, Y.; Zhang, D.; Zhang, S.; Jia, Q.; Zhang, H. Quality Confirmation of Electrical Measurement Data for Space Parts Based on MSA Method. In *Signal and Information Processing, Networking and Computers, Proceedings of the 7th International Conference on Signal and Information Processing, Networking and Computers (ICSINC), Rizhao, China, 21–23 September 2020*; Wang, Y., Fu, M., Xu, L., Zou, J., Eds.; Springer: Singapore, 2020; pp. 344–351.
37. Paneva, V.I. Evaluation of Adequacy of Quantitative Laboratory Chemical Analysis Techniques. *Inorg. Mater.* **2009**, *45*, 1652–1657. [CrossRef]

38. Cosmi, F.; Dal Maso, A. BES TEST^{TM} Accuracy Evaluation by Means of 3D-Printed Phantoms. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2023**, *237*, 1783–1792. [CrossRef]

39. Garcia, O.; Di Giorgio, M.; Vallerga, M.B.; Radl, A.; Taja, M.R.; Seoane, A.; De Luca, J.; Stuck Oliveira, M.; Valdivia, P.; Lamadrid, A.I. Interlaboratory Comparison of Dicentric Chromosome Assay Using Electronically Transmitted Images. *Radiat. Prot. Dosim.* **2013**, *154*, 18–25. [CrossRef] [PubMed]

40. 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* **2022**, *187*, 110340. [CrossRef]

41. Bich, W.; Callegaro, L.; Pennecchi, F. Non-Linear Models and Best Estimates in the GUM. *Metrologia* **2006**, *43*, S196. [CrossRef]

42. Carmona, J.A.; Ramírez, P.; García, M.C.; Santos, J.; Muñoz, J. Linear and Non-Linear Flow Behavior of Welan Gum Solutions. *Rheol. Acta* **2019**, *58*, 1–8. [CrossRef]

43. Bodnar, O.; Wübbeler, G.; Elster, C. On the Application of Supplement 1 to the GUM to Non-Linear Problems. *Metrologia* **2011**, *48*, 333. [CrossRef]

44. Cox, M.; Harris, P.; Siebert, B.R.-L. Evaluation of Measurement Uncertainty Based on the Propagation of Distributions Using Monte Carlo Simulation. *Meas. Tech.* **2003**, *46*, 824–833. [CrossRef]

45. Chen, A.; Chen, C. Comparison of GUM and Monte Carlo Methods for Evaluating Measurement Uncertainty of Perspiration Measurement Systems. *Measurement* **2016**, *87*, 27–37. [CrossRef]

46. Mahmoud, G.M.; Hegazy, R.S. Comparison of GUM and Monte Carlo Methods for the Uncertainty Estimation in Hardness Measurements. *Int. J. Metrol. Qual. Eng.* **2017**, *8*, 14. [CrossRef]

47. Niemeier, W.; Tengen, D. Uncertainty Assessment in Geodetic Network Adjustment by Combining GUM and Monte-Carlo-Simulations. *J. Appl. Geod.* **2017**, *11*, 67–76. [CrossRef]

48. Papadopoulos, C.E.; Yeung, H. Uncertainty Estimation and Monte Carlo Simulation Method. *Flow Meas. Instrum.* **2001**, *12*, 291–298. [CrossRef]

49. Garg, N.; Yadav, S.; Aswal, D.K. Monte Carlo Simulation in Uncertainty Evaluation: Strategy, Implications and Future Prospects. *MAPAN* **2019**, *34*, 299–304. [CrossRef]

50. Wang, L.; Xiao, T.; Liu, S.; Zhang, W.; Yang, B.; Chen, L. Quantification of Model Uncertainty and Variability for Landslide Displacement Prediction Based on Monte Carlo Simulation. *Gondwana Res.* **2023**, *123*, 27–40. [CrossRef]

51. Possolo, A.; Merkatas, C.; Bodnar, O. Asymmetrical Uncertainties. *Metrologia* **2019**, *56*, 45009. [CrossRef]

52. Honarvar, A. Modeling of Asymmetry between Gasoline and Crude Oil Prices: A Monte Carlo Comparison. *Comput. Econ.* **2010**, *36*, 237–262. [CrossRef]

53. Huang, H. A Propensity-Based Framework for Measurement Uncertainty Analysis. *Measurement* **2023**, *213*, 112693. [CrossRef]

54. Rashki, M. The Soft Monte Carlo Method. *Appl. Math. Model.* **2021**, *94*, 558–575. [CrossRef]

55. Sengupta, A.; Mondal, S.; Das, A.; Guler, S.I. A Bayesian Approach to Quantifying Uncertainties and Improving Generalizability in Traffic Prediction Models. *Transp. Res. Part C Emerg. Technol.* **2024**, *162*, 104585. [CrossRef]

56. Meija, J.; Bodnar, O.; Possolo, A. Ode to Bayesian Methods in Metrology. *Metrologia* **2023**, *60*, 52001. [CrossRef]

57. Bardsley, J.M. MCMC-Based Image Reconstruction with Uncertainty Quantification. *SIAM J. Sci. Comput.* **2012**, *34*, A1316–A1332. [CrossRef]

58. Green, P.L.; Worden, K. Bayesian and Markov Chain Monte Carlo Methods for Identifying Nonlinear Systems in the Presence of Uncertainty. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *373*, 20140405. [CrossRef] [PubMed]

59. Chandra, R.; Simmons, J. Bayesian Neural Networks via MCMC: A Python-Based Tutorial. *IEEE Access* **2024**, *12*, 70519–70549. [CrossRef]

60. Carobbi, C.; Pennecchi, F. Bayesian Conformity Assessment in Presence of Systematic Measurement Errors. *Metrologia* **2016**, *53*, S74. [CrossRef]

61. Elster, C. Calculation of Uncertainty in the Presence of Prior Knowledge. *Metrologia* **2007**, *44*, 111–116. [CrossRef]

62. Stephan, K.E.; Penny, W.D.; Daunizeau, J.; Moran, R.J.; Friston, K.J. Bayesian Model Selection for Group Studies. *NeuroImage* **2009**, *46*, 1004–1017. [CrossRef] [PubMed]

63. van de Schoot, R.; Depaoli, S.; King, R.; Kramer, B.; Märtens, K.; Tadesse, M.G.; Vannucci, M.; Gelman, A.; Veen, D.; Willemsen, J.; et al. Bayesian Statistics and Modelling. *Nat. Rev. Methods Primers* **2021**, *1*, 1–26. [CrossRef]

64. Demeyer, S.; Fischer, N.; Elster, C. Guidance on Bayesian Uncertainty Evaluation for a Class of GUM Measurement Models. *Metrologia* **2020**, *58*, 14001. [CrossRef]

65. Abduljaleel, M.; Midi, H.; Karimi, M. Outlier Detection in the Analysis of Nested Gage R&R, Random Effect Model. *Stat. Transit. New Ser.* **2019**, *20*, 31–56. [CrossRef]

66. Al-Qudah, S.K. A Study of the AIAG Measurement System Analysis (MSA) Method for Quality Control. *J. Manag. Eng. Integr.* **2017**, *10*, 68–80.

67. Al-Refaie, A.; Bata, N. Evaluating Measurement and Process Capabilities by GR&R with Four Quality Measures. *Measurement* **2010**, *43*, 842–851. [CrossRef]

68. Aquila, G.; Peruchi, R.S.; Junior, P.R.; Rocha, L.C.S.; de Queiroz, A.R.; de Oliveira Pamplona, E.; Balestrassi, P.P. Analysis of the Wind Average Speed in Different Brazilian States Using the Nested GR&R Measurement System. *Measurement* **2018**, *115*, 217–222. [CrossRef]

69. Cepova, L.; Kovacikova, A.; Cep, R.; Klaput, P.; Mizera, O. Measurement System Analyses–Gauge Repeatability and Reproducibility Methods. *Meas. Sci. Rev.* **2018**, *18*, 20–27. [CrossRef]

70. Kooshan, F. Implementation of Measurement System Analysis System (MSA): In the Piston Ring Company Case Study. *Int. J. Sci. Technol.* **2012**, *2*, 749–761.

71. Marques, R.A.M.; Pereira, R.B.D.; Peruchi, R.S.; Brandao, L.C.; Ferreira, J.R.; Davim, J.P. Multivariate GR&R through Factor Analysis. *Measurement* **2020**, *151*, 107107. [CrossRef]

72. Peruchi, R.S.; Balestrassi, P.P.; de Paiva, A.P.; Ferreira, J.R.; de Santana Carmelossi, M. A New Multivariate Gage R&R Method for Correlated Characteristics. *Int. J. Prod. Econ.* **2013**, *144*, 301–315. [CrossRef]

73. Razumić, A.; Runje, B.; Keran, Z.; Trzun, Z.; Pugar, D. Reproducibility of Areal Topography Parameters Obtained by Atomic Force Microscope. *Teh. Glas.* **2025**, *19*, 1–6. [CrossRef]

74. Bontempi, M.; Cardinale, U.; Bragonzoni, L.; Muccioli, G.M.M.; Alesi, D.; di Matteo, B.; Marcacci, M.; Zaffagnini, S. A Computer Simulation Protocol to Assess the Accuracy of a Radio Stereometric Analysis (RSA) Image Processor According to the ISO-5725. *arXiv* **2020**, arXiv:2006.03913.

75. D'Aucelli, G.M.; Giaquinto, N.; Mannatrizio, S.; Savino, M. A Fully Customizable MATLAB Framework for MSA Based on ISO 5725 Standard. *Meas. Sci. Technol.* **2017**, *28*, 44007. [CrossRef]

76. Stöckl, D.; Van Uytfanghe, K.; Cabaleiro, D.; Thienpont, L. Calculation of Measurement Uncertainty in Clinical Chemistry. *Clin. Chem.* **2005**, *51*, 276–277. [CrossRef] [PubMed]

77. Feinberg, M. Basics of Interlaboratory Studies: The Trends in the New ISO 5725 Standard Edition. *TrAC Trends Anal. Chem.* **1995**, *14*, 450–457. [CrossRef]

78. Razumić, A. Procjena Mjerne Nesigurnosti Rezultata Mjerenja na Području Mikroskopije Atomskih Sila u Dimenzijskom Nanomjeriteljstvu. Ph.D. Thesis, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Zagreb, Croatia, 2024.

79. Alar, V.; Razumić, A.; Runje, B.; Stojanović, I.; Kurtela, M.; Štrbac, B. Application of Areal Topography Parameters in Surface Characterization. *Appl. Sci.* **2025**, *15*, 6573. [CrossRef]

80. Razumić, A.; Runje, B.; Lisjak, D.; Kolar, D.; Horvatić Novak, A.; Štrbac, B.; Savković, B. Atomic Force Microscopy: Step Height Measurement Uncertainty Evaluation. *Teh. Glas.* **2024**, *18*, 209–214. [CrossRef]

81. van der Veen, A.; Cox, M.; Greenwood, J.; Bošnjaković, A.; Karahodžić, V.; Martens, S.; Klauenberg, K.; Elster, C.; Demeyer, S.; Fischer, N.; et al. *Good Practice in Evaluating Measurement Uncertainty—Compendium of Examples*; EURAMET: Teddington, UK, 2021.

82. Cox, M.G.; Greenwood, J.; Bošnjaković, A.; Karahodžić, V. Two-point and Multipoint Calibration. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 57–68.

83. Ribeiro, A.S.; Cox, M.G.; Almeida, M.C.; Sousa, J.A.; Martins, L.L.; Simoes, C.; Brito, R.; Loureiro, D.; Silva, M.A.; Soares, A.C. Evaluation of Measurement Uncertainty in Average Areal Rainfall—Uncertainty Propagation for Three Methods. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 205–218.

84. Martins, L.L.; Ribeiro, A.S.; Cox, M.G.; Sousa, J.A.; Loureiro, D.; Almeida, M.C.; Silva, M.A.; Brito, R.; Soares, A.C. Evaluation of Measurement Uncertainty in SBI—Single Burning Item Reaction to Fire Test. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 77–90.

85. Cox, M.G.; Gardiner, T.; Robinson, R.; Smith, T.; Ellison, S.L.R.; Van der Veen, A.M.H. Greenhouse Gas Emission Inventories. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 269–282.

86. Demeyer, S.; Fischer, N.; Cox, M.G.; Van der Veen, A.M.H.; Sousa, J.A.; Pellegrino, O.; Bošnjaković, A.; Karahodžić, V.; Elster, C. Bayesian Approach Applied to the Mass Calibration Example in JCGM 101:2008. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 71–76.

87. Caebergs, T.; De Boeck, B.; Pétry, J.; Sebaïhi, N.; Cox, M.G.; Fischer, N.; Greenwood, J. Uncertainty Evaluation of Nanoparticle Size by AFM, by Means of an Optimised Design of Experiment for a Hierarchical Mixed Model in a Bayesian Framework Approach. In *Good Practice in Evaluating Measurement Uncertainty*; EURAMET: Teddington, UK, 2021; pp. 171–188.

88. *ISO 15189:2022*; Medical Laboratories—Requirements for Quality and Competence. International Organization for Standardization: Geneva, Switzerland, 2022.

89. *ISO/IEC 17020:2012*; Conformity Assessment—Requirements for the Operation of Various Types of Bodies Performing Inspection. International Organization for Standardization: Geneva, Switzerland, 2012.

90. *ISO 10012:2003*; Measurement Management Systems—Requirements for Measurement Processes and Measuring Equipment. International Organization for Standardization: Geneva, Switzerland, 2003.

91. *ISO 14253-1:2017*; Geometrical Product Specifications (GPS)—Inspection by Measurement of Workpieces and Measuring Equipment Part 1: Decision Rules for Verifying Conformity or Nonconformity with Specifications. International Organization for Standardization: Geneva, Switzerland, 2017.

92. *ISO 9001:2015*; Quality Management Systems—Requirements. International Organization for Standardization: Geneva, Switzerland, 2015.

93. Keran, Z.; Runje, B.; Piljek, P.; Razumić, A. Roboforming in ISF—Characteristics, Development, and the Step Towards Industry 5.0. *Sustainability* **2025**, *17*, 2562. [CrossRef]

94. Shankar, R.; Gupta, L. Modelling Risks in Transition from Industry 4.0 to Industry 5.0. *Ann. Oper. Res.* **2024**, *342*, 1275–1320. [CrossRef]

95. Zio, E.; Guarnieri, F. Industry 5.0: Do Risk Assessment and Risk Management Need to Update? And If Yes, How? *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2025**, *239*, 389–396. [CrossRef]

96. Begoli, E.; Bhattacharya, T.; Kusnezov, D. The Need for Uncertainty Quantification in Machine-Assisted Medical Decision Making. *Nat. Mach. Intell.* **2019**, *1*, 20–23. [CrossRef]

97. Nagel, C.; Pilia, N.; Loewe, A.; Dössel, O. Quantification of Interpatient 12-Lead ECG Variabilities within a Healthy Cohort. *Curr. Dir. Biomed. Eng.* **2020**, *6*, 493–496. [CrossRef]

98. Aston, P.J.; Mehari, T.; Bosnjakovic, A.; Harris, P.M.; Sundar, A.; Williams, S.E.; Dössel, O.; Loewe, A.; Nagel, C.; Strodthoff, N. Multi-Class ECG Feature Importance Rankings: Cardiologists vs Algorithms. In Proceedings of the 2022 Computing in Cardiology (CinC), Tampere, Finland, 4–7 September 2022; Volume 498, pp. 1–4.

99. Mehari, T.; Sundar, A.; Bosnjakovic, A.; Harris, P.; Williams, S.E.; Loewe, A.; Doessel, O.; Nagel, C.; Strodthoff, N.; Aston, P.J. Ecg Feature Importance Rankings: Cardiologists vs. Algorithms. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 2014–2024. [CrossRef] [PubMed]