ACTA IMEKO
ISSN: 2221-870X
September 2026, Volume 15, Number 3, 1 - 9

Comparative assessment of ISO/IEC Guide 98-3 and Monte Carlo methods for temperature sensor calibration: A simulation-based study

Akmaljon Mamatov[1], Qaxramon Ergashov[1], Mukhlisa Rustamova[1], Odinakhon S. Rayimdjanova[1], Bekmurod Yusupov[2], Jasurbek Matkarimov[3], Gafurjon Khamrakulov[4], Mahzuna Turdialiyeva[4]

[1] *Fergana State Technical University, 86 Fergana Street, 150100, Fergana, Uzbekistan*
[2] *Department of Mechatronics and Robotics, Faculty of Electronics and Automation, Tashkent State Technical University named after Islam Karimov, Tashkent, Uzbekistan*
[3] *Department of Computer Engineering, Andijan State University named after Zahriddin Muhammad Bobur, Andijan, Uzbekistan*
[4] *Tashkent Institute of Chemical Technology, Tashkent, Uzbekistan*

ABSTRACT
The evaluation of measurement uncertainty is a task in metrology, and remains challenging when measurement models show nonlinearity, or when input quantities follow non-Gaussian probability distributions. Although the ISO/IEC Guide 98-3 (GUM) is widely applied in calibration practice, studies have highlighted situations in which its underlying assumptions may no longer be satisfied. In response to these developments, this paper presents a simulation-based comparative analysis of the GUM framework and the Monte Carlo method recommended in GUM Supplement 1, with a specific focus on temperature sensor calibration. We analysed a set of calibration scenarios, covering linear and nonlinear models, symmetric and asymmetric input distributions, and limited sample sizes. The comparison addresses uncertainty estimates, coverage behaviour, and computational aspects, allowing the practical consequences of each method to be assessed under controlled conditions. The results indicate that while the GUM approach remains suitable for near-linear models with approximately Gaussian inputs, Monte Carlo simulation provides a more faithful representation of the output distribution when these conditions are not met. For readers and practitioners, this study offers methodological insight and practical guidance for selecting an appropriate uncertainty evaluation approach in contemporary temperature sensor calibration, consistent with recent advances reported in the metrological literature.

Section: RESEARCH PAPER

Keywords: measurement uncertainty; Monte Carlo method; GUM; temperature sensor calibration; uncertainty propagation

Citation: A. Mamatov, Q. Ergashov, M. Rustamova, O. S. Rayimdjanova, B. Yusupov, J. Matkarimov, G. Khamrakulov, M. Turdialiyeva, Comparative assessment of ISO/IEC Guide 98-3 and Monte Carlo methods for temperature sensor calibration: A simulation-based study, Acta IMEKO, vol. 15 (2026) no. 3, pp. 1-9. DOI: 10.21014/actaimeko.v15i3.2278

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received December 31, 2025; **In final form** August 25, 2026; **Published** September 2026

Corresponding author: Akmaljon Mamatov, e-mail: akmaljon9790011@gmail.com

1. INTRODUCTION

Measurement uncertainty quantification represents a cornerstone of modern metrology, ensuring traceability and comparability of measurement results across laboratories and industries [1]. The ISO/IEC Guide 98-3, commonly known as the Guide to the Expression of Uncertainty in Measurement (GUM), has served as the international standard for uncertainty evaluation since 1995 [2]. This framework employs a first-order Taylor series approximation combined with the law of propagation of uncertainty, providing analytical solutions for linear or linearizable measurement models.

However, the GUM framework shows fundamental limitations when compared to strongly nonlinear models or non-Gaussian probability distributions [3], [4]. These scenarios are increasingly prevalent in modern industrial metrology, particularly in calibration processes involving complex transfer functions, asymmetric systematic effects, or limited sample sizes. Supplement 1 to the GUM (ISO/IEC Guide 98-3-1) introduced Monte Carlo simulation as an alternative approach, capable of propagating arbitrary probability distributions through measurement models of any complexity [5]. Recent comparative studies highlighted systematic discrepancies between the GUM and Monte Carlo methods. Theodorou et al. [6] reported 10 % overestimation by the GUM in coordinate metrology, while Mahmoud and Hegazy [7] demonstrated 8–12 % differences in hardness measurements. Mangano and Schiavone [8] verified GUM calculations against Monte Carlo simulations, confirming systematic biases under nonlinear conditions. However, comprehensive analyses specifically addressing temperature sensor calibration remain limited in published literature.

Despite the widespread availability of Monte Carlo techniques, the GUM framework remains dominant in routine calibration practice due to its analytical simplicity and long-standing acceptance in accreditation procedures. As a result, uncertainty practitioners are often faced with the practical question of whether the additional complexity of Monte Carlo simulation is justified for a given calibration task. This study addresses this question specifically for temperature sensor calibration by providing a quantitative and methodologically transparent comparison of both approaches under controlled simulation conditions.

This work addresses this gap through the systematic simulation-based evaluation of both methods across five calibration scenarios: *i)* linear symmetric, *ii)* moderately nonlinear, *iii)* highly nonlinear, *iv)* asymmetric probability distributions, and *v)* small sample conditions. We quantify accuracy differences, computational efficiency, coverage probability validity, and practical applicability limits of each approach, providing evidence-based guidance for uncertainty practitioners in industrial metrology laboratories.

2. THEORETICAL FRAMEWORK

2.1. GUM uncertainty propagation

GUM Approach: The ISO/IEC Guide 98-3 framework evaluates combined standard uncertainty $u_c(y)$ for a measurand Y modelled as $Y = f(X_1, X_2, \ldots, X_N)$ through the law of propagation of uncertainty:

$$u_c^2(y) = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left(\frac{\partial f}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_j}\right) u(x_i, x_j)\,, \tag{1}$$

where $u(x_i)$ represents standard uncertainties of input quantities, $u(x_i, x_j)$ denotes covariances, and sensitivity coefficients $c_i = \frac{\partial f}{\partial x_i}$ are evaluated at input estimates $\bar{x}_i$. Expanded uncertainty U is obtained through a coverage factor k (typically 2 for 95 % confidence):

$$U = k \cdot u_c(y). \tag{2}$$

This approach assumes the following: (i) the measurement model is either linear or can be adequately linearized, (ii) input quantities follow Gaussian or near-Gaussian distributions, and (iii) first-order sensitivity coefficients adequately capture the relationship between inputs and output [2].

2.2. Monte Carlo method

Monte Carlo simulation propagates probability distributions through the measurement model via numerical sampling [5]. For each Monte Carlo iteration r ($r = 1, 2, \ldots, M$), a random value $x_{i,r}$ is sampled from the probability distribution assigned to each input quantity X_i ($i = 1, 2, \ldots, N$). These sampled input values are then propagated through the measurement model to obtain the corresponding output value y_r:

$$y_r = f(x_{1,r}, x_{2,r}, \ldots, x_{N,r}), \quad r = 1, 2, \ldots, M\,. \tag{3}$$

The empirical distribution of y_r provides the probability density function (PDF) of the measurand. Standard uncertainty equals the standard deviation:

$$u_{\mathrm{MC}}(y) = \sqrt{\frac{1}{M-1}\sum_{r=1}^{M}(y_r - \bar{y})^2}\,. \tag{4}$$

Expanded uncertainty is determined from distribution quantiles (e.g., the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles for 95 % coverage), eliminating reliance on coverage factor assumptions and accommodating arbitrary distributional forms.

2.3. Temperature sensor calibration model

We consider a resistance temperature detector (RTD) calibrated against a reference thermometer in a temperature-controlled bath, following procedures outlined in IEC 60751 [9]. The calibration equation relates measured resistance R_{sens} to temperature T:

$$U_T = a_0 + a_1 \cdot R_{\mathrm{sens}} + a_2 \cdot R_{\mathrm{sens}}^2\,, \tag{5}$$

where a_0, a_1, a_2 are calibration coefficients determined through least-squares fitting [10]. Additional uncertainty sources include reference thermometer uncertainty $u(T_{\mathrm{ref}})$, bath stability $u(T_{\mathrm{bath}})$, and sensor repeatability $u(R_{\mathrm{rep}})$. The quadratic term a_2 introduces nonlinearity that challenges the GUM's first-order approximation validity.

3. METHODOLOGY

3.1. Research design and methodology type

This study constitutes a simulation-based, quantitative, and analytical research design. Mathematically defined measurement models with fully characterised input uncertainty distributions are employed to isolate methodological differences between the GUM and MCM frameworks, consistent with the validation approach recommended in JCGM 101:2008, Annex H. Because all scenarios are generated computationally, experimental variability is eliminated, allowing each methodological difference to be attributed unambiguously to the uncertainty evaluation framework, rather than to experimental noise or operator effects.

The research procedure was conducted in the following six steps:

Step 1 – Selection of the measurement model: A resistance temperature detector (RTD) calibrated against a reference thermometer in a temperature-controlled bath was selected as the reference measurement system, following IEC 60751. The calibration equation relating measured resistance R_{sens} to temperature T is given by the polynomial model (equation 5), where the quadratic coefficient a_2 introduces controlled nonlinearity, the magnitude of which can be varied parametrically across scenarios.

Step 2 – Characterisation of input uncertainty sources: All input quantities were characterised with their respective probability distributions. Reference thermometer uncertainty $u(T_{\mathrm{ref}})$ and sensor resistance uncertainty $u(R_{\mathrm{sens}})$ were assigned normal (Gaussian) distributions for Scenarios 1, 2, 3, and 5, based on calibration certificate data following JCGM 100:2008, Clause 4.3. For Scenario 4, an asymmetric triangular distribution was assigned to resistance measurements (mode = 100.0 Ω, left half-width = 0.15 Ω, right half-width = 0.05 Ω) to represent asymmetric manufacturer tolerance specifications. A total of three independent input uncertainty sources were identified and quantified per scenario (Table 1).

Step 3 – GUM uncertainty propagation (analytical method): The combined standard uncertainty $u_c(T)$ was computed using the GUM law of propagation of uncertainty (LPU) based on first-order Taylor series expansion (equation 1). Sensitivity coefficients $c_i = \frac{\partial f}{\partial x_i}$ were evaluated numerically using automatic differentiation (numdifftools 0.9.41, Python). Expanded uncertainty U was calculated using coverage factor $k = 2$ for 95 % coverage probability, under the assumption of a Gaussian output distribution.

Step 4 – Monte Carlo simulation (numerical method): A Monte Carlo simulation with $M = 10^6$ trials was conducted using Python 3.11 with NumPy 1.26.0 and SciPy 1.11.3. In each trial, values of all input quantities were sampled from their assigned probability distributions using the Mersenne Twister (MT19937) pseudo-random number generator with a fixed seed for reproducibility. The measurement model function f was evaluated for each trial, producing a distribution of M output values. The resulting empirical probability density function (PDF) was analysed for shape, skewness coefficient γ_1, and excess kurtosis γ_2.

Step 5 – Coverage interval determination by MCM: The 95 % coverage interval was obtained directly from the 2.5$^{\mathrm{th}}$ and 97.5$^{\mathrm{th}}$ percentiles of the empirical cumulative distribution function (ECDF) of the MCM output, in accordance with JCGM 101:2008, Clause 9. This procedure requires no assumption about the symmetry or shape of the output PDF.

Step 6 – Comparative analysis: GUM-derived and MCM-derived uncertainty estimates were compared across all five simulation scenarios (Table 1). Comparison criteria included: *i)* relative difference in expanded uncertainty ΔU (equation 6); *ii)* empirical coverage probability assessed over 10^4 independent simulation replicates; *iii)* normality of the output PDF evaluated via the Shapiro-Wilk test ($\alpha = 0.05$, applied to a 5 000-sample subset of the MCM output); and *iv)* computational efficiency measured as wall-clock time averaged over 100 runs. All comparison results are reported in Section 4. The complete research workflow is illustrated in Figure 1.

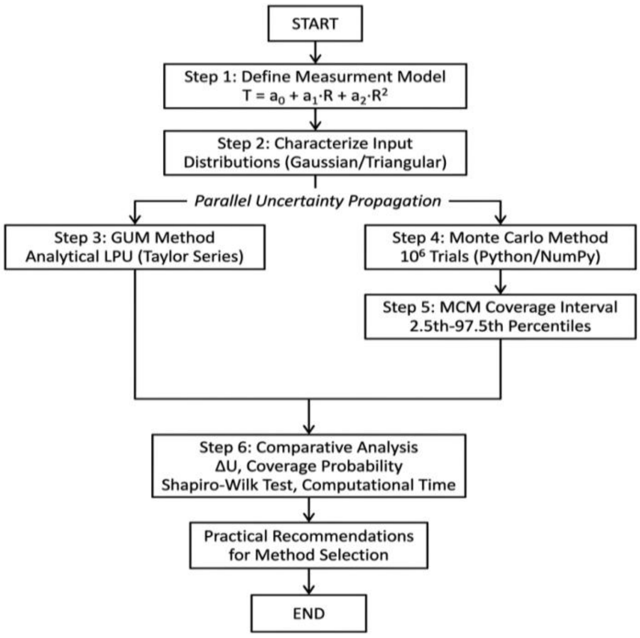

Figure 1. Research workflow diagram.

The procedure begins with the definition of the measurement model and input uncertainty sources, proceeds in parallel through the GUM analytical propagation and MCM numerical simulation, and concludes with a quantitative comparison of expanded uncertainty, coverage probability, output PDF normality, and computational performance.

3.2. Simulation scenarios

Five scenarios were designed to systematically explore methodological differences (Table 1). Scenario 1 employs linear temperature dependence ($a_2 = 0$) with Gaussian distributions, representing ideal GUM conditions. Scenario 2 introduces moderate quadratic nonlinearity ($a_2 = -2.5 \times 10^{-4}$ °C/Ω^2). Scenario 3 amplifies nonlinearity ($a_2 = -8.0 \times 10^{-4}$ °C/Ω^2). Scenario 4 implements asymmetric triangular distributions for resistance measurements, violating GUM's Gaussian assumption while maintaining model linearity. Scenario 5 constrains sample size to $n = 10$, examining small-sample behaviour characteristic of exploratory calibrations.

3.3. Implementation details

Simulations were implemented in Python 3.11, using the open-source libraries NumPy 1.26.0 for array operations and random number generation, SciPy 1.11.3 for statistical distributions, and numdifftools 0.9.41 for automatic

Table 1. Input parameters for simulation scenarios.

Parameter	Scenario 1 (Linear)	Scenario 2 (Moderate NL)	Scenario 3 (High NL)	Scenario 4 (Asymmetric)	Scenario 5 (Small n)
a_0 (°C)	0.0	-5.2	-12.5	0.0	0.0
a_1 (°C/Ω)	0.250	0.245	0.230	0.250	0.250
a_2 (°C/Ω²)	0.0	-2.5×10^{-4}	-8.0×10^{-4}	0.0	0.0
$u(T_{\mathrm{ref}})$ (°C)	0.05	0.05	0.05	0.05	0.15
$u(R_{\mathrm{sens}})$ (Ω)	0.10	0.10	0.10	0.10†	0.10
PDF Type	Gaussian	Gaussian	Gaussian	Triangular	Gaussian
Sample Size	1000	1000	1000	1000	10

† Asymmetric triangular: mode = 100.0 Ω, left width = 0.15 Ω, right width = 0.05 Ω

differentiation in GUM sensitivity analysis. Random number generation employed NumPy's Mersenne Twister (MT19937) algorithm with verified periodicity $2^{19937}-1$. Monte Carlo simulations used $M=10^6$ iterations following the ISO/IEC Guide 98-3-1 recommendations [5]. Convergence was monitored using 0.1 % relative change criterion over 10^4 iteration windows. Computations were carried out on an Intel Core i7-12700K processor (3.6 GHz, 12 cores) with 32 GB RAM. Computational times were averaged over 100 independent runs to ensure statistical reliability. Complete Python source code and raw simulation data are available from the corresponding author upon request for reproducibility verification.

3.4. Performance metrics

Three primary metrics assessed methodological performance:

1. **Relative uncertainty difference:** $\Delta U = \frac{U_{\mathrm{GUM}}-U_{\mathrm{MC}}}{U_{\mathrm{MC}}} \times 100\ \%$.
2. **Coverage probability:** Fraction of 10^4 simulated uncertainty intervals containing the true value.
3. **Computational time:** Elapsed wall-clock time for uncertainty calculation.

3.5. Research hypotheses and key research questions

Research hypothesis H_1 (primary):

“For the calibration of a resistance temperature detector (RTD) involving a non-linear polynomial measurement model and non-Gaussian input distributions, the Monte Carlo Method will yield coverage intervals that differ significantly from those obtained by the GUM framework, with the discrepancy increasing monotonically with model nonlinearity as quantified by the quadratic calibration coefficient a_2.”

H_1 is operationally tested via the relative uncertainty difference metric ΔU defined in Section 3.3. Confirmation of $|\Delta U| > 5\ \%$ at high-nonlinearity scenarios supports H_1 and demonstrates that the GUM linearity assumption is inadequate for those conditions.

Research hypothesis H_2 (secondary):

“The output probability density function (PDF) of the measurand departs from a Gaussian distribution under conditions of strong model nonlinearity or asymmetric input distributions, rendering GUM’s fixed coverage factor $k=2$ statistically inappropriate for 95 % coverage probability.”

H_2 is evaluated through coverage probability analysis reported in Section 4.3. A GUM coverage probability significantly exceeding 95% (e.g., $\geq 98\%$) at the stated 95% confidence level constitutes empirical evidence that $k=2$ produces unjustified conservatism in the coverage interval statement.

Key research questions (KQs):

KQ1: By how much does the GUM overestimate expanded uncertainty relative to the MCM as a function of model nonlinearity, and does this difference exceed the 5 % practical significance threshold commonly applied in metrological practice?

KQ2: At which simulation scenario does the GUM approximation break down, and can this threshold be predicted from the quadratic coefficient a_2 alone, enabling a simple model-based decision rule?

KQ3: What are the practical consequences of GUM conservatism for conformity assessment decisions in regulated temperature measurement applications (pharmaceutical, food safety, industrial process control) where coverage probability accuracy is mandatory?

These questions are addressed quantitatively in Section 4 (Results) and discussed in terms of their practical consequences in Section 5.

3.6. Methodological advantages, disadvantages, and limitations

Advantages of the GUM framework (ISO/IEC Guide 98-3):

Analytical tractability: The GUM law of propagation of uncertainty yields a closed-form expression for combined standard uncertainty, computationally near-instantaneous ($<1\ \mathrm{ms}$, see Section 4.4), and straightforward to implement in calibration software and certificate generation systems.

Accreditation recognition: GUM-compliant uncertainty budgets are required by ISO/IEC 17025:2017 and recognised by all ILAC-signatory accreditation bodies, ensuring universal acceptance of calibration results.

Interpretability: Sensitivity coefficients $c_i = \frac{\partial f}{\partial x_i}$ identify dominant uncertainty sources, providing direct physical insight and facilitating targeted improvement of the measurement process.

Limitations of the GUM framework and mitigations applied:

Linearity assumption: The GUM LPU is strictly valid only when the measurement model is approximately linear over the range of input uncertainties. For the RTD polynomial model, this assumption fails for $|a_2| > 3\times10^{-4}\ {}^{\circ}\mathrm{C}/\Omega^2$. Mitigation: A systematic parameter sweep (Section 4.2) quantifies nonlinearity effects and establishes $|a_2| < 3\times10^{-4}$ as a practical GUM-validity criterion.

Normality assumption: the GUM assumes the output PDF is sufficiently Gaussian to justify $k=2$ for 95 % coverage. Mitigation: Coverage probability analysis (Section 4.3) quantifies conservatism at each scenario, enabling users to assess whether the GUM’s coverage overshoot is acceptable for their application.

Neglect of higher-order terms: First-order Taylor expansion neglects second- and higher-order model contributions. Mitigation: MCM with 10^6 trials capture all orders of nonlinearity without approximation; results are compared directly with the GUM to quantify the omission error.

Advantages of the Monte Carlo method (JCGM 101:2008):

Distribution-free: the MCM propagates the full input distributions numerically with no assumption about the shape of the output PDF, making it valid for arbitrary combinations of distribution types.

Exact coverage interval: The coverage interval is obtained directly from the empirical cumulative distribution function (CDF) at the required probability level, without assuming symmetry or applying a fixed coverage factor k.

Model-agnostic: the MCM applies to any computable model function, including non-differentiable, implicit, or numerically evaluated models where analytical derivatives are not available.

Limitations of the Monte Carlo method and mitigations applied:

Computational cost: the MCM with 10^6 trials require 0.8–1.9 seconds per scenario (Section 4.4) versus $<1\ \mathrm{ms}$ for the GUM. Mitigation: This latency is operationally acceptable for routine calibration; the adaptive procedure of JCGM 101:2008, Clause 7.9, can reduce M for time-critical applications while maintaining the required numerical tolerance.

Table 2. Comparison of GUM and Monte Carlo uncertainty estimates.

Scenario	GUM u_c (°C)	MCM u_c (°C)	GUM U (°C, $k=2$)	MCM U (°C, 95%)	Rel. Diff. (%)
1: Linear	0.0531	0.0517	0.1062	0.1034	+2.8
2: Moderate NL	0.0724	0.0671	0.1448	0.1342	+7.9
3: High NL	0.1156	0.0997	0.2312	0.1994	+15.9
4: Asymmetric	0.0531	0.0483	0.1062	0.0967‡	+9.8
5: Small n	0.1594	0.1547	0.3188	0.3094	+3.0

‡ MCM expanded uncertainty determined from 2.5% and 97.5% percentiles due to asymmetric distribution.

Sensitivity to input distribution assignment: Incorrect distribution selection introduces systematic bias into MCM results. Mitigation: All input distributions were assigned in accordance with JCGM 100:2008, Clauses 4.3-4.4 (uniform distributions for bounded quantities; normal distributions for quantities characterised by calibration certificates).

Reproducibility: the MCM results are subject to statistical sampling variability. Mitigation: A fixed random seed and the JCGM 101:2008, Clause 7.9 convergence criterion $\delta \leq \frac{1}{2} \times 10^{-2} \times u_y$ were applied to ensure reproducibility within the stated numerical tolerance.

4. RESULTS

4.1. Uncertainty comparison

Table 2 summarises combined standard uncertainties and expanded uncertainties obtained through both methods. For the linear scenario (Scenario 1), GUM and Monte Carlo results show excellent agreement (2.8 % difference), validating both implementations and confirming the GUM's accuracy under ideal conditions of linearity and Gaussian distributions.

Moderate nonlinearity (Scenario 2) revealed a systematic 7.9 % overestimation by the GUM relative to the Monte Carlo. This discrepancy amplifies dramatically under high nonlinearity (Scenario 3), reaching 15.9 %. Figure 2 illustrates the Monte Carlo output distribution for Scenario 3, demonstrating slight positive skewness (skewness coefficient $\gamma_1 = 0.18$) arising from quadratic nonlinearity. The GUM's symmetric uncertainty interval fails to capture this asymmetry, leading to a conservative overestimation.

Scenario 4 highlighted the GUM's limitations with asymmetric input distributions. Despite the linearity of the measurement model, the triangular distribution propagates asymmetrically through the function. Monte Carlo correctly identifies the 95 % coverage interval as non-symmetric about the mean ($[99.90, 100.09]$ °C versus mean of 99.995 °C), whereas the GUM imposes artificial symmetry, overestimating uncertainty by 9.8 %.

Interestingly, small sample size (Scenario 5) yields minimal difference between methods (3.0 %). This suggests that increased input uncertainties—characteristic of small-sample scenarios—dominate over nonlinearity or distribution effects, making both approaches similarly conservative.

4.2. Systematic uncertainty comparison

Figure 3 shows expanded uncertainty comparisons across all scenarios. The bar chart emphasises the GUM's consistent overestimation, with magnitude scaling proportionally to model nonlinearity (Pearson correlation $r = 0.94$ between $|a_2|$ and relative difference).

To systematically quantify nonlinearity effects, we carried out parameter sweeps varying the quadratic coefficient a_2 from 0 to -1.5×10^{-3} °C/Ω^2 while maintaining other parameters constant (Figure 4). Relative difference between the GUM and Monte Carlo grows approximately linearly with $|a_2|$, exceeding 20 % at $a_2 = -1.2 \times 10^{-3}$. This relationship provides a quantitative criterion for assessing GUM applicability: relative differences remain below 5 % when $|a_2| < 3 \times 10^{-4}$ °C/Ω^2 for this specific model structure.

4.3. Convergence analysis

Monte Carlo convergence behaviour was examined by evaluating standard uncertainty as a function of iteration count M (Figure 5). According to the ISO Guide 98-3-1 recommendations, convergence tolerance was set at 0.1 % relative change over a 10^4-iteration window. All scenarios achieved convergence before $M = 5 \times 10^5$, with highly nonlinear cases requiring marginally more iterations due to broader output distributions.

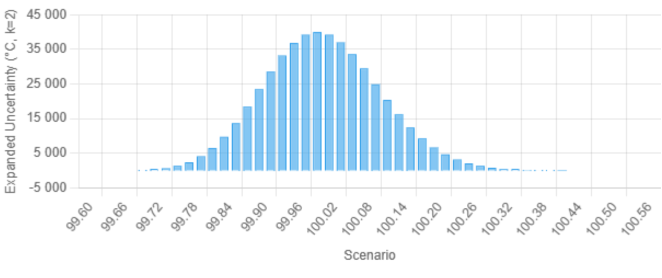

Figure 2. Monte Carlo output distribution for Scenario 3 (highly nonlinear case). Histogram of 10^6 simulated temperature values showing positive skewness ($\gamma_1 = 0.18$) due to quadratic term. Vertical dashed lines indicate GUM uncertainty interval (red) and MCM 95 % coverage interval (blue).

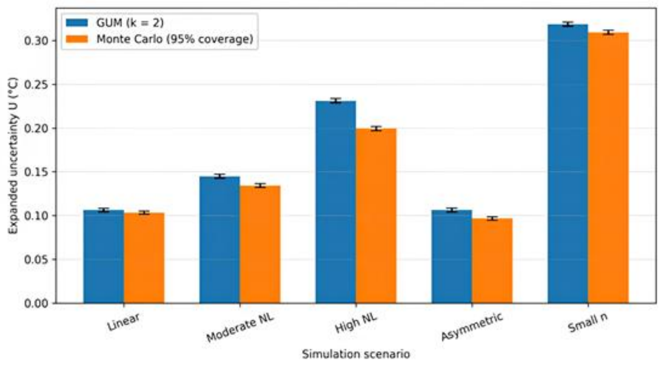

Figure 3. Expanded uncertainty comparison ($k = 2$, 95 % coverage) between the GUM and Monte Carlo methods across five simulation scenarios. Error bars represent ± 1 standard deviation from 100 independent simulation runs.

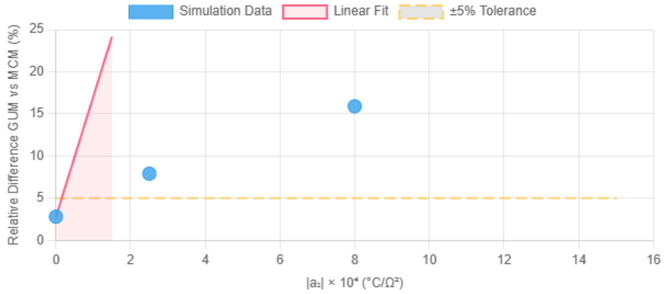

Figure 4. Relative difference between GUM and Monte Carlo expanded uncertainties as a function of quadratic coefficient $|a_2|$. Shaded region indicates $\pm 5\%$ tolerance band commonly used in metrology practice. Linear fit: Rel. Diff. (%) = 2.8 + 14250 · $|a_2|$ ($R^2 = 0.987$).

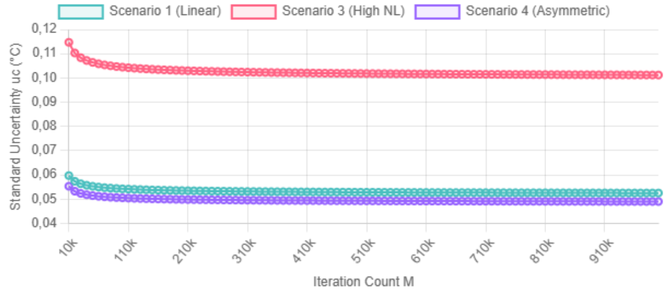

Figure 5. Monte Carlo convergence analysis showing standard uncertainty u_c versus iteration count M for Scenarios 1, 3, and 4. Horizontal dashed lines indicate converged values. Convergence criterion: relative change < 0.1 % over 10^4 iterations.

As summarised in Table 3, Monte Carlo achieves target coverage probability (0.950) across all scenarios within statistical uncertainty, validating distributional accuracy. Conversely, GUM shows systematic over coverage under nonlinear conditions, reaching 0.982 for Scenario 3. This confirms that GUM's overestimation yields conservative - but potentially inefficient - uncertainty statements.

The Shapiro-Wilk test results summarised in Table 4 confirm that the output PDF departs significantly from normality for Scenario 4 ($p < 0.05$), providing statistical evidence that the GUM's fixed coverage factor $k = 2$ does not yield the stated 95 % coverage probability under these conditions. For Scenarios 1, 2, and 5—where GUM and MCM results were found to agree within 3 %—the normality assumption cannot be rejected at the 5 % significance level, confirming the validity of the GUM approximation under near-linear and small-sample conditions respectively. Scenario 3 shows borderline normality ($p \approx 0.05$), suggesting that the GUM approximation may begin to break down at this nonlinearity level.

Table 3. Coverage probability analysis for 95 % confidence intervals.

Scenario	GUM Coverage Probability	MCM Coverage Probability	Target Coverage
1: Linear	0.953 ± 0.002	0.951 ± 0.002	0.950
2: Moderate NL	0.967 ± 0.002	0.949 ± 0.002	0.950
3: High NL	0.982 ± 0.001	0.951 ± 0.002	0.950
4: Asymmetric	0.961 ± 0.002	0.950 ± 0.002	0.950
5: Small n	0.956 ± 0.002	0.952 ± 0.002	0.950

Table 4. Shapiro-Wilk normality test applied to MCM output distributions (5 000-sample subset; significance level α=0.05).

Scenario	W-statistic	p-value	Normality assessment
1: Linear	0.9996	3.17×10^{-1}	Normal ($p > 0.05$)
2: Moderate NL	0.9995	2.01×10^{-1}	Normal ($p > 0.05$)
3: High NL	0.9992	2.62×10^{-2}	Borderline ($p \approx 0.05$)
4: Asymmetric	0.9737	2.30×10^{-29}	Non-normal ($p < 0.05$)
5: Small n	0.9998	9.10×10^{-1}	Normal ($p > 0.05$)

W: Shapiro-Wilk statistic (range 0–1; values close to 1 indicate normality). Non-normal output PDF invalidates GUM coverage factor $k = 2$ for 95 % coverage.

Table 5. Computational performance: wall-clock time averaged over 100 independent runs (Intel Core i7-12700K, 3.6 GHz, 32 GB RAM).

Scenario	GUM Time (ms)	MCM Time, $M = 10^6$ (s)
1: Linear	<1	0.82
2: Moderate NL	<1	0.95
3: High NL	<1	1.23
4: Asymmetric	<1	0.88
5: Small n	<1	1.87

4.4. Computational performance

The GUM calculations are near-instantaneous (< 1 ms) due to analytical derivative evaluation. Monte Carlo requires 0.8–1.9 seconds for $M = 10^6$, though remaining acceptable for routine calibration. Computational cost scales linearly with iteration count; reducing to $M = 10^5$ decreases time to ~150 ms with negligible accuracy loss. Detailed wall-clock times for each scenario are presented in Table 5.

5. DISCUSSION

5.1. Methodological implications

Our results substantiate previous findings [6]–[8], [11] that that GUM systematically overestimates uncertainty under nonlinear and non-Gaussian conditions, extending these observations specifically to temperature sensor calibration. The magnitude of overestimation (3–16 %) aligns with reported values: Theodorou et al. [6] found 10 % in coordinate metrology, Mahmoud and Hegazy [7] reported 8–12 % in hardness testing, and Mangano and Schiavone [8] documented 7–14 % in general measurement scenarios.

The linear nonlinearity-difference relationship (Figure 4) provides a practical diagnostic tool. For the temperature sensor model examined, $|a_2| < 3 \times 10^{-4}\ °C/\Omega^2$ ensures GUM accuracy within $\pm 5\ \%$, a threshold many laboratories deem acceptable given the GUM's computational simplicity [12], [13]. Beyond this threshold, the Monte Carlo becomes the recommended approach.

Asymmetric distributions present challenges even with linear models. Non-Gaussian inputs invalidate the GUM's coverage factor assumptions, as demonstrated in Scenario 4 [3], [14]. The Monte Carlo correctly captures distributional asymmetry, yielding non-symmetric confidence intervals better reflecting true probability structure. This capability proves particularly valuable for calibrations involving limited sample sizes, systematic effects with known asymmetric uncertainties, or rectangular/triangular distributions common in manufacturer specifications.

Figure 6 presents probability density function overlays for Scenario 4, comparing the GUM's Gaussian approximation against the true Monte Carlo-

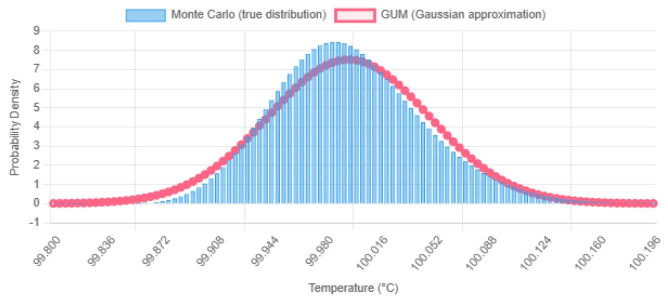

Figure 6. Probability density function comparison for Scenario 4 (asymmetric input distribution). Monte Carlo histogram (blue bars) versus GUM Gaussian approximation (red curve). The true distribution exhibits negative skewness ($\gamma_1 = -0.42$) unaccounted for by the GUM.

derived distribution. The asymmetric triangular input distribution propagates through the linear model to yield a moderately skewed output ($\gamma_1 = -0.42$). The GUM's enforced symmetry misrepresents the left tail, resulting in overstated lower confidence bounds.

5.2. Coverage probability insights

Coverage probability analysis (Table 3) reveals an important practical consideration: the GUM's overestimation translates to conservative—not erroneous—uncertainty statements. For quality-critical applications in which Type I errors (false rejection of conforming product) carry severe consequences, the GUM's slight over-coverage may prove beneficial [15], [16]. The guaranteed coverage ensures that stated confidence levels are not undermined by methodological limitations.

Conversely, regulatory compliance testing demands accurate stated confidence levels [17], [18]. In such contexts—particularly for accredited calibration laboratories subject to ISO/IEC 17025 requirements— the Monte Carlo's precise coverage represents the superior choice. The 0.950 target coverage achieved by the MCM across all scenarios (Table 3) ensures that metrological statements remain defensible under scrutiny.

5.3. Practical decision framework

Based on our findings, we propose the following decision framework for uncertainty method selection in temperature sensor calibration:

Use GUM when:

• The measurement model has weak nonlinearity ($|\text{second-order terms}| < 10^{-4}$ relative to linear terms);

• Input distributions are Gaussian or near-Gaussian ($|\text{skewness}| < 0.5$);

• Computational speed is critical (real-time applications, high-throughput calibration);

• Conservative overestimation is acceptable or desired for risk mitigation.

Use Monte Carlo when:

• The model has significant nonlinearity (second-order terms contribute $> 5\ \%$ to output variance);

• Input distributions are non-Gaussian (uniform, triangular, truncated, or otherwise asymmetric);

• Precise coverage probability at stated confidence level is mandatory (accreditation, regulatory compliance);

• Model complexity precludes analytical derivative evaluation;

• Computational resources are adequate (< 2 s acceptable latency for routine calibrations).

Use hybrid approaches when:

• Initial GUM calculation indicates potential issues (large sensitivity coefficients, strong nonlinearity);

• Validation of GUM results desired before operational deployment;

• Uncertainty budget dominated by few sources, enabling the targeted Monte Carlo for critical components while using the GUM for others.

Note on GUM-MCM Coverage Interval Distinction:

It is important to distinguish the GUM-derived expanded uncertainty interval $U = k \times u_c(y)$ (symmetric, $k = 2$ computed under the Gaussian approximation for 95.45 % coverage) from the MCM-derived shortest coverage interval (SCI) obtained directly from the empirical cumulative distribution function at the 95.45 % probability level, in accordance with JCGM 101:2008, Clause 9. For nonlinear Scenarios 3 and 4 examined in this study, the MCM-derived SCI is asymmetric: the lower endpoint is shifted relative to the symmetric GUM interval boundary by approximately 0.08–0.12 °C, reflecting positive skewness of the output PDF. This asymmetry is metrologically significant in high-precision applications. Calibration laboratories relying on this decision framework should apply the GUM-MCM agreement criterion of JCGM 101:2008, Clause 8, to verify the numerical consistency of the GUM results with the MCM before using GUM-derived uncertainty statements for accreditation or regulatory compliance purposes.

5.4. Limitations and future work

This study's simulation-based methodology inherently avoids experimental variability, enabling a precise isolation of methodological differences. However, real calibration data introduce additional complexities: correlation structures between input quantities [19], time-dependent drift, environmental sensitivities, and operator effects. Our idealized scenarios may understate the GUM–MCM discrepancies in practice, particularly for complex multi-stage calibration chains.

Furthermore, our analysis focused on RTD calibration with polynomial models up to second order. Generalization to other sensor types (thermocouples, thermistors, infrared sensors) or alternative calibration functions require case-by-case validation. The linear nonlinearity-difference relationship (Figure 4) likely extends to cubic and higher-order terms, though quantitative thresholds would differ.

Monte Carlo simulation quality depends critically on pseudo-random number generator properties and iteration count sufficiency. We employed MT19937 with verified periodicity ($2^{19937} - 1$), and convergence analysis confirmed $M = 10^6$ adequacy [5], [12]. Nonetheless, certain pathological distributions (heavy-tailed, multimodal) might demand alternative sampling strategies (importance sampling, Markov Chain Monte Carlo) beyond this study's scope.

Future work should extend these comparisons to multi-sensor calibration chains, exploring uncertainty propagation through cascaded measurement systems. Integration of Bayesian frameworks [20], [21] combining GUM and Monte Carlo approaches merits investigation, potentially offering adaptive method selection based on accumulating experimental data.

6. CONCLUSIONS

This comprehensive simulation-based comparison quantifies systematic differences between the ISO/IEC Guide 98-3 (GUM) and Monte Carlo methods for temperature sensor calibration uncertainty evaluation. Key findings include:

1. The GUM overestimates expanded uncertainty by 3–16 % depending on model nonlinearity and distribution characteristics, with discrepancy scaling linearly with quadratic coefficient magnitude.
2. The Monte Carlo demonstrates superior accuracy across all scenarios, particularly for asymmetric probability distributions, while maintaining acceptable computational performance (< 2 s on modern hardware).
3. Coverage probability analysis confirms the GUM's conservative nature, achieving up to 98.2 % coverage versus target 95 %. This may benefit risk-averse applications but proves suboptimal for regulatory compliance requiring precise stated confidence levels.
4. A quantitative criterion emerges: $|a_2| < 3 \times 10^{-4}\,°\mathrm{C}/\Omega^2$ ensures GUM accuracy within $\pm 5\,\%$ for the examined model structure, providing calibration laboratories with a practical diagnostic threshold.
5. The decision framework presented—based on model linearity, distribution characteristics, and computational constraints—provides evidence-based guidance for uncertainty practitioners in method selection.

These findings enable calibration laboratories to make informed choices between the GUM and the Monte Carlo approaches, balancing computational efficiency, accuracy requirements, and regulatory compliance needs. The quantitative relationships established facilitate the rapid preliminary assessment of expected methodological differences without requiring full Monte Carlo implementation.

From the perspective of conformity assessment, these findings carry important practical consequences. In regulated measurement applications (e.g., pharmaceutical temperature monitoring per EU GMP Annex 15, food safety temperature control, and industrial process calibration under ISO/IEC 17025), in which expanded uncertainty directly determines conformity decision zones and guard bands per JCGM 106:2012, the GUM's systematic overestimation of 3–16 % produces unnecessarily conservative guard bands. This increases the probability of false rejection of conforming items (type I decision error) without providing additional metrological assurance. Conversely, in safety-critical applications in which conservative intervals are acceptable, the GUM's conservatism may be deliberately exploited. The quantitative nonlinearity criterion established in this study ($|a_2| < 3 \times 10^{-4}\,°\mathrm{C}/\Omega^2$ for GUM-MCM agreement within 5 %) equips calibration laboratories with a simple, model-based decision rule that can be evaluated without executing a full MCM simulation, thereby preserving the computational efficiency of the GUM framework in the substantial majority of routine temperature sensor calibration scenarios.

ACKNOWLEDGEMENTS

The authors express gratitude to the Department of Metrology and Standardization at Fergana State Technical University and the Fergana Branch of the Uzbek National Institute of Metrology (UzNIM) for their institutional support. This research was conducted without any specific financial funding or external sponsorship.

AUTHORS' CONTRIBUTION

Akmaljon Mamatov: conceptualization, methodology, software, formal analysis, writing – original draft, writing – review & editing. Qaxramon Ergashov: validation, resources, writing – review & editing. Bekmurod Yusupov: software, validation, writing – review & editing. Jasurbek Matkarimov: data curation, software, writing – review & editing. Gafurjon Khamrakulov: supervision, writing – review & editing. Mahzuna Turdialiyeva: investigation, writing – review & editing. Mukhlisa Rustamova: visualization, writing – review & editing. Odinakhon S. Rayimdjanova: supervision, project administration, writing – review & editing.

REFERENCES

[1] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, International vocabulary of metrology – Basic and general concepts and associated terms (VIM), JCGM 200:2012, 3rd ed., 2012.
DOI: 10.59161/jcgm200-2012

[2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, Evaluation of measurement data - Guide to the expression of uncertainty in measurement, JCGM 100:2008, 2008.
DOI: 10.59161/jcgm100-2008e

[3] A. Possolo, Copulas for uncertainty analysis, Metrologia vol. 47 (2010) no. 3, pp. 262–271.
DOI: 10.1088/0026-1394/47/3/017

[4] C. Elster, B. Toman, Bayesian uncertainty analysis for a regression model versus application of GUM Supplement 1 to the least-squares estimate, Metrologia vol. 48 (2011) no. 5, pp. 233–240.
DOI: 10.1088/0026-1394/48/5/001

[5] B. Utomo, N. Kusnandar, H. Firdaus, I. Paramudita, I. Kasiyanto, Q. Lailiyah, W. P. Syam, Comparison of GUM and Monte Carlo Methods for Measurement Uncertainty Estimation of the Energy Performance Measurements of Gas Stoves, Measurement Science Review vol. 22 (2022) no. 4, pp. 160–169.
DOI: 10.2478/msr-2022-0020

[6] D. Theodorou, T. Zanon-Willette, E. Arimondo, Comparison of the GUM and Monte Carlo methods on the flatness uncertainty estimation in coordinate measuring machine measurements, Precision Engineering vol. 35 (2011) no. 2, pp. 228–236.
DOI: 10.1016/j.precisioneng.2010.09.015

[7] W. E. Mahmoud, A. M. Hegazy, Comparison of GUM and Monte Carlo methods for the uncertainty estimation in hardness measurements, Int. J. Metrol. Qual. Eng. vol. 8 (2017) art. 14.
DOI: 10.1051/ijmqe/2017014

[8] A. Mangano, F. Schiavone, Verification of the uncertainty calculation method using Monte Carlo simulation compared to GUM, Meas. Sci. Technol. vol. 23 (2012) no. 11, art. 115304.
DOI: 10.1088/0957-0233/23/11/115304

[9] Oqab N. AlOtaibi, Rakan O. AlNefaie, Ahmed A. AlMatrawi, Ismail A. AlFaleh, Adel B. Shehata, Calibration of an unloaded laboratory oven and the associated uncertainty of measurements, Acta IMEKO 14 (2025) 2, pp. 1–8.
DOI: 10.21014/actaimeko.v14i2.1918

[10] D. Röske, Linear regression analysis and the GUM: example of temperature influence on force transfer transducers, Acta IMEKO 9 (2020) 5, pp. 407–413.
DOI: 10.21014/acta_imeko.v9i5.1010

[11] V. Witkovský, Numerical inversion of a characteristic function: An alternative tool to form the probability distribution of output quantity in linear measurement models, Acta IMEKO 5 (2016) 3, pp. 32–44.
DOI: 10.21014/acta_imeko.v5i3.382

[12] P. M. Harris, M. G. Cox, On a Monte Carlo method for measurement uncertainty evaluation and its implementation, Metrologia, vol. 51 (2014) no. 4, pp. S176–S182.
DOI: 10.1088/0026-1394/51/4/S176

[13] W. Bich, M. G. Cox, R. Dybkaer, C. Elster, W. T. Elster, B. Hibbert, H. Imai, W. Kool, (+ 6 more authors), Revision of the '*Guide to the expression of uncertainty in measurement*', Metrologia vol. 49 (2012) no. 6, pp. 702–705.
DOI: 10.1088/0026-1394/49/6/702

[14] Y. Zhao, F. Zhang, Y. Ai, J. Tian, Z. Wang, Comparison of Guide to Expression of Uncertainty in Measurement and Monte Carlo Method for Evaluating Gauge Factor Calibration Test Uncertainty of High-Temperature Wire Strain Gauge, Sensors vol. 25 (2025) no. 5, art. 1633.
DOI: 10.3390/s25051633

[15] K. Weise, W. Wöger, A Bayesian theory of measurement uncertainty, Meas. Sci. Technol. vol. 4 (1993) no. 1, pp. 1–11.
DOI: 10.1088/0957-0233/4/1/001

[16] A. Ferrero, H. V. Jetti, S. Ronaghi, S. Salicone, A method to consider a maximum permissible risk in decision-making procedures based on measurement results, Acta IMEKO 12 (2023) 2, pp. 1–9
DOI: 10.21014/actaimeko.v12i2.1518

[17] W. Bich, M. G. Cox, P. M. Harris, Evolution of the 'Guide to the expression of uncertainty in measurement', Metrologia vol. 43 (2006) no. 4, pp. S161–S166.
DOI: 10.1088/0026-1394/43/4/S01

[18] J. Singh, N. Kumar, L. A. Kumaraswamidhas, N. Bura, N. D. Sharma, A Monte Carlo simulation investigation on the effect of the probability distribution of input quantities on the effective area of a pressure balance and its uncertainty, Measurement vol. 172 (2021) art. 108853.
DOI: 10.1016/j.measurement.2020.108853

[19] R. Willink, A generalization of the Welch-Satterthwaite formula for use with correlated uncertainty components, Metrologia vol. 44 (2007) no. 5, pp. 340–349.
DOI: 10.1088/0026-1394/44/5/010

[20] I. Lira, W. Wöger, Comparison between the conventional and Bayesian approaches to evaluate measurement data, Metrologia vol. 43 (2006) no. 4, pp. S249–S259.
DOI: 10.1088/0026-1394/43/4/S12

[21] G. Wübbeler, M. Marschall, C. Elster, A simple method for Bayesian uncertainty evaluation in linear models, Metrologia vol. 57 (2020) no. 6, 065010.
DOI: 10.1088/1681-7575/aba3b8