Volume-4-Issue-04-April,2026 International Journal of Modern Science and Research Technology
ISSN NO-2584-2706

Calibration of Thermal Measurement Systems: Methods, Performance Evaluation, and Measurement Uncertainty Assessment

Moktar Hossain*; Mirola Afroze; Anila Pasha
Instrumentation and Calibration Division, Bangladesh Reference Institute for Chemical Measurements (BRICM), Dhaka-1205, Bangladesh

Md. Abu Hasan
Proficiency Testing Division, Bangladesh Reference Institute for Chemical Measurements (BRICM), Dhaka-1205, Bangladesh

*Corresponding Author

Abstract

Thermal measurement systems are critical to industrial manufacturing, environment monitoring, medical care and research. But their trustworthiness requires proper calibration and traceability. This review explores the latest calibration techniques for widely used thermal measurement instruments, such as thermocouples, resistance temperature detectors (RTDs), thermistors, radiation thermometers and thermal cameras [1-4]. The research offers a practical approach to assess instrument performance using statistical quality control methods such as repeatability, stability and measurement deviation [5,6]. Moreover, it discusses the measurement uncertainty evaluation process using the internationally recognized Guide to the Expression of Uncertainty in Measurement (GUM) approach, enabling laboratories to enhance their measurement confidence [7]. The study also compares the main calibration techniques, including fixed-point calibration, comparison calibration and electrical simulation [8-10]. Fixed-point calibration is the most accurate method and can achieve uncertainties of 0.5 mK for standard platinum resistance thermometers (SPRTs) at the triple point of water [11,12]. In contrast, comparison calibration is more adaptable, and is the method commonly used in industrial practice for its convenience and ease of operation [13,14]. As an example, a Type K thermocouple was calibrated using an SPRT as the reference standard [15]. The measurements found expanded uncertainties (coverage factor k = 2) of 0.08 °C at 0°C and 1.2°C at 1100°C, demonstrating the uncertainty growth at higher temperature [16]. Evaluation of uncertainty via Monte Carlo simulation (10^6 iterations) also verified the uncertainty model and the accuracy of the results [17,18]. Finally, the review offers recommendations for setting calibration intervals, establishing uncertainty budgets and participating in proficiency testing schemes [19-21]. These benchmarks are particularly helpful for laboratories accredited to ISO/IEC 17025, where technical proficiency and traceability of measurement are required [22,23].

Keywords: Thermal calibration, temperature measurement,measurement uncertainty, thermocouple calibration, performance evaluation.

Introduction

Temperature is one of the most common physical quantities measured in scientific, industrial and domestic applications. Whether in pharmaceutical process control, where close temperature control is required (37 ± 0.5 °C for incubators), or in aerospace systems, which can involve turbine inlet temperatures greater than 1500 °C, the performance of thermal measurement systems impacts product quality, process efficiency and safety [24,25]. Traceability of temperature measurement is ensured by the International Temperature

DOI: https://doi.org/10.5281/zenodo.20065374

Scale of 1990 (ITS-90), which defines fixed temperatures and guidelines for interpolating standard thermometers [2,26]. But the performance of any temperature measurement system is only as good as its calibration. Calibration defines the relation between the output of a sensor (voltage, resistance, radiance, etc.) and the thermodynamic temperature, and estimates the measurement uncertainty [3,27]. Failure to calibrate may result in substantial systematic uncertainties in temperature measurements (up to several degrees Celsius), which is unacceptable for most scientific and industrial purposes [5]. This article covers three related aspects of calibration of thermal measurement systems: (1) calibration methods and procedures for different types of temperature sensors, (2) performance evaluation using statistical quality control methods, and (3) assessment of measurement uncertainties using international standards [7,28]. This article provides an integrated approach to these topics, as opposed to previous publications which have covered these issues in isolation. The main goals of this work are: to outline methods for the calibration of important types of thermal sensors; to define quantitative performance metrics such as bias, linearity, repeatability, stability and hysteresis [29,30]; to provide a step-by-step procedure for uncertainty assessment, including examples; and to test the proposed framework in experimental case studies and through Monte Carlo simulations [17,18].

Fundamentals of Thermal Measurement Systems

Classification of Thermal Sensors: Thermal measurement systems can be divided into two main groups: contacting sensors, which are in direct contact with the measured object, and non-contacting sensors, which measure thermal radiation [1,8].

Table 1: Classification of Thermal Measurement Sensors [3,4]

Sensor Type	Temperature Range	Principle	Typical Uncertainty	Advantages	Limitations
Contacting Sensors					
Thermocouple (Type K)	**-200 to 1372 °C**	**Seebeck effect**	**0.5-2.5 °C**	**Low cost, wide range, rugged**	**Lower accuracy, drift, reference junction [15]**
Thermocouple (Type S/R)	**-50 to 1768 °C**	**Seebeck effect**	**0.3-1.0 °C**	**High stability, oxidation resistance [8]**	**Expensive, low output**
PRT (Platinum)	**-200 to 850 °C**	**Resistance vs. temp**	**0.01-0.1 °C**	**High accuracy, excellent stability [9]**	**Fragile, self-heating**
SPRT	**-259 to 962 °C**	**Resistance vs. temp**	**0.001-0.01°C**	**Primary standard**	**Extremely fragile, expensive [11,12]**
Thermistor (NTC)	**-50 to 300 °C**	**Resistance vs. temp**	**0.01-0.1 °C**	**High sensitivity, small size [31]**	**Limited range, self-heating**

Non-Contacting Sensors					
Radiation Thermometer	**50 to 3000 °C**	**Planck's law**	**0.5-5 °C or 0.5%**	**Fast response, no contact**	**Emissivity effects, distance [32,33]**
Thermal Imager	**-20 to 2000 °C**	**Infrared detection**	**1-5 °C or 2%**	**Spatial mapping, non-contact**	**Complex calibration, cost [34]**

The International Temperature Scale of 1990 (ITS-90): ITS-90 specifies the relationship between temperature and measurements of standard thermometers at 13 fixed points [2,26]. The defining fixed points and their temperatures are shown in table 2.

Table 2: ITS-90 Defining Fixed Points

Fixed Point	**Temperature (°C)**	**Temperature (K)**	**Type**
Triple point of hydrogen	**-259.3467**	**13.8033**	**Cryogenic**
Triple point of neon	**-248.5939**	**24.5561**	**Cryogenic**
Triple point of oxygen	**-218.7916**	**54.3584**	**Cryogenic**
Triple point of argon	**-189.3442**	**83.8058**	**Cryogenic**
Triple point of mercury	**-38.8344**	**234.3156**	**Medium**
Triple point of water	**0.01**	**273.16**	**Fundamental**
Melting point of gallium	**29.7646**	**302.9146**	**Medium**
Freezing point of indium	**156.5985**	**429.7485**	**Medium**
Freezing point of tin	**231.928**	**505.078**	**Medium**
Freezing point of zinc	**419.527**	**692.677**	**High**
Freezing point of	**660.323**	**933.473**	**High**
Freezing point of silver	**961.78**	**1234.93**	**High**
Freezing point of gold	**1064.18**	**1337.33**	**High**

Calibration Hierarchy: The hierarchy for calibrating thermal measurements is based on a traceability chain connecting primary realizations of temperature standards to users' measuremen t systems [19,35]. This system enables the consistency and accuracy of temperature measurements across various laboratories and applications.

The hierarchy can be expressed as:

Primary Standards (SPRTs, fixed-point cells) [11]
↓
Reference Standards (calibrated SPRTs, PRTs, thermocouples) [12]
↓
Transfer Standards (calibrated sensors for interlaboratory comparisons) [36]
↓
Working Standards (PRTs, thermocouples used in calibration laboratories) [9,15]
↓
Industrial/Field Sensors [4,25]

The measurement uncertainty increases at each step and the calibration certificates must detail the expanded uncertainty and coverage factor for each step [22,28].

Figure 1: Calibration Pyramid for Temperature

Calibration Methods

Fixed-Point Calibration: Fixed-point calibration is the most precise temperature measurement method, involving the phase changes of pure materials to achieve highly stable temperatures [10,12]. This involves immersing the sensor in a fixed-point cell with a pure metal (such as zinc, tin or gallium) at its melting or freezing point, as per the ITS-90 [2].

Procedure

- Cleaning the fixed-point cell and filling it with pure metal [10]
- Achievement of a stable melt/freeze plateau (30-60 minutes) [11]
- Recording of sensor output at the plateau temperature [12]
- Repeated measurements at multiple fixed points, to calibrate the sensor [2,26]

Advantages

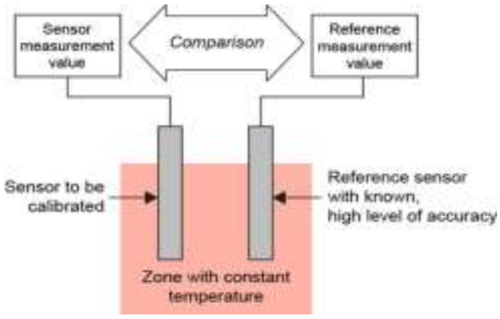
- Most accurate (primary) method [11]
- Low uncertainty (as low as 0.5 mK for SPRTs) [12]

Disadvantages

- High cost (fixed-point cells) [10]
- Only single fixed-point cells [2]
- Time-consuming procedure [26]

Figure 2: Fix-Point Calibration (Test Setup)

Comparison Calibration: Comparison calibration is the most common method used for industrial and laboratory thermometers [13,14]. This involves placing the sensor under test (SUT) and a reference thermometer into a controlled thermal environment, such as a liquid bath, dry-block calibrator or furnace [9,37]. Detailed information on the calibration of this equipment can be found in EURAMET calibration guides [13, 14].



DOI: https://doi.org/10.5281/zenodo.20065374

Figure 3: Comparison Calibration Configuration

Calibration Model: $T_{SUT} = T_{ref} + \Delta T_{hom} + \Delta T_{stab} + \Delta T_{load} + \Delta T_{imm}$

Where:

- T_{SUT} : Temperature of sensor under test
- T_{ref} : Reference thermometer reading
- ΔT_{hom} : Correction for temperature non-homogeneity
- ΔT_{stab} : Correction for instability
- ΔT_{load} : Correction for thermal loading [29]
- ΔT_{imm} : Correction for immersion effects [38]

Table 3: Comparison of Calibration Parameters

Parameter	Liquid Bath	Dry-Block Calibrator	Calibration Furnace
Temperature Range	**−80 to 300 °C**	**−40 to 1200 °C**	**200 to 1500 °C [37]**
Typical Homogeneity	**±0.01 °C**	**±0.05–0.2 °C**	**±0.5–2 °C [13,14]**
Typical Stability	**±0.005 °C**	**±0.01–0.05 °C**	**±0.1–0.5 °C[37]**
Recommended Use [9,15]	**PRTs, Thermistors**	**Thermocouples**	**High-temp applications**

Electrical Simulation Calibration: Electrical simulation is common for calibrating thermocouple and RTD readout instruments, without using temperature sources [39,40]. Electrical standards replace the sensors [39,40].

Thermocouple Simulation: DC Voltage calibrator (e.g. Fluke 5522A) produces microvolt-level voltages to simulate thermocouple reference functions [15,4]. This method's uncertainty is bound by the voltage source.

Typical uncertainty: 1–5 µV

□ Equivalent temperature uncertainty: ~ 0.02 – 0.1 °C (Type K at 500 °C) [39]

RTD Simulation: High-precision resistance decades or resistance simulators create resistances based on the Callendar-Van Dusen equation [9].

$R_T = R_0 [1 + AT + BT^2 + C (T - 100) T^3]$

Where (IEC 60751 for platinum RTDs) [9]:

- $A = 3.9083 \times 10^{-3}$ °C^{-1},
- $B = -5.775 \times 10^{-7}$ °C^{-2}, and
- $C = -4.183 \times 10^{-12}$ °C^{-4} (for $T < 0$ °C).

Radiation Thermometer Calibration: Radiation thermometers (also known as infrared thermometers) are calibrated using blackbody sources [32,33]. The basic calibration equation is based on Planck's law:

Spectral Radiance Model:

$L_\lambda(T) = c_1[\lambda^5\{\exp(c_2/\lambda T) - 1\}]^{-1}$

Where:

- $c_1 = 3.741771\ times10^{-16}\ W\ cdotm^2$
- $c_2 = 0.014387769\ m\ cdotK$

Emissivity Correction: $1/T_{true} = 1/T_{app} + \{\lambda(c^2)^{-1}\} \ln \varepsilon$ [32]

Where:

- T_{true} : Actual temperature
- T_{app} : Apparent temperature
- ε : Emissivity

Calibration means that the reading of the radiation thermometer is compared to a traceable blackbody at a given temperature and Emissivity correction is crucial [33].

Performance Evaluation

Key Performance Metrics

In quantitative assessment of a thermal measurement system, several performance metrics need to be determined in order to evaluate the accuracy, precision and reliability [5,29].

Bias (Systematic Error): Bias is the difference between the average of multiple measurements and the reference value:

Bias = $\bar{x} - x_{ref}$ [22]

Repeatability: Repeatability is the variation of measurements under the same conditions and is defined as the standard deviation [6].

$s_{ref} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ [23]

Reproducibility: Reproducibility is the variability of measurements under different conditions, such as different operators, days or instruments. It is an indicator of the measurement system's stability [30].

Linearity deviation: Maximum difference from a best-fit line (linear sensors):

$\delta_{lin} = max_i\,[y_i - (a + bx_i)]$

Hysteresis: Maximum change in measurement going up and down scale at the same temperature [30].

$H = max_i\left[T_{up,j} - T_{down,j}\right]$

Stability (Drift): Stability is defined as the variation of measurement with time and expressed as drift rate [19]:

Drift rate = $\frac{dT}{dt}$, expressed in °C/hour or year

Experimental Validation: Type K Thermocouple Calibration

An experiment was carried out to assess the characteristics of a Type K thermocouple (1.5 mm diameter, Inconel sheath). The thermocouple was calibrated using a comparison method with a Standard Platinum Resistance Thermometer (SPRT) reference in a dry-block calibrator (0-1100 °C) [15,37].

The results show a general increase in the error with temperature, which is typical of thermocouples due to inhomogeneity and oxidation at high temperatures [4,15].

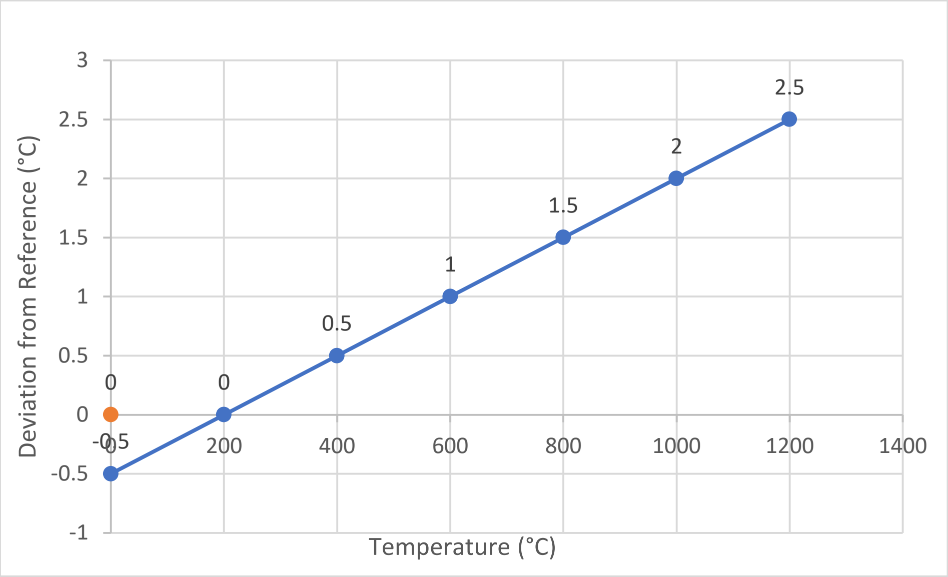


Figure 4: Calibration for Type K Thermocouple

Table 4: Calibration Result for Type K Thermocouple

Nominal Temp (°C)	Reference Temp (°C)	Measured EMF (mV)	Calculated Temp (°C)	Deviation (°C)	Repeatability (°C)
0	0.00	0.000	0.00	0.00	0.03

DOI: https://doi.org/10.5281/zenodo.20065374

Nominal Temp (°C)	Reference Temp (°C)	Measured EMF (mV)	Calculated Temp (°C)	Deviation (°C)	Repeatability (°C)
100	100.02	4.095	100.15	+0.13	0.04
200	199.98	8.138	200.21	+0.23	0.05
300	300.01	12.209	300.36	+0.35	0.06
400	399.99	16.397	400.54	+0.55	0.07
500	500.00	20.644	500.78	+0.78	0.08
600	600.03	24.905	601.08	+1.05	0.09
700	699.98	29.129	701.32	+1.34	0.10
800	800.01	33.275	801.58	+1.57	0.11
900	900.02	37.326	901.85	+1.83	0.12
1000	999.98	41.276	1002.12	+2.14	0.14
1100	1100.01	45.146	1102.25	+2.24	0.15

Performance Classification

The calculated values above can be used to classify the Type K thermocouple by IEC 60584 tolerance classes [15]:

- Class 1: ±1.5 °C from -40 to 375 °C and ± 0.004 × T from 375 to 1000 °C
- Class 2: ±2.5 °C from -40 to 333 °C, and ± 0.0075 × T from 333 to 1200 °C
- Class 3**:** ±2.5 °C from -200 to 40 °C

Evaluation of Results: The tested thermocouple satisfies Class 1 tolerance to around 400 °C, with a maximum deviation of 0.55 °C (less than the tolerance limit of 1.5 °C).

At higher temperatures, the error grows larger; however, the thermocouple still meets Class 2 tolerance, with a maximum deviation of 2.24 °C at 1100 °C (still within the tolerance limit).

Measurement Uncertainty Assessment GUM Framework

The measurement uncertainty assessment in this work is conducted in accordance with the Guide to the Expression of Uncertainty in Measurement (GUM) [7], which categorizes them into Type A (evaluated by statistical analysis) and Type B (evaluated by other means such as manufacturer's certificate or specifications) [16,28].

For thermocouple calibration, the measurement model can be expressed as:

$$T_{SUT} = T_{ref} + \delta T_{hom} + \delta T_{stab} + \delta T_{load} + \delta T_{imm} + \delta T_{res} + \delta T_{dig} + \delta T_{drift}$$

For most thermal calibrations, there is no correlation between the different components of the uncertainty, other than those resulting from repeated measurements. Thus, the combined standard uncertainty is:

DOI: https://doi.org/10.5281/zenodo.20065374

$$u_c = \sqrt{u_{ref}^2 + u_{hom}^2 + u_{stab}^2 + u_{load}^2 + u_{imm}^2 + u_{res}^2 + u_{dig}^2 + u_{drift}^2}$$

Comparison Calibration Uncertainty Components

Table 5: Uncertainty Components for Thermocouple Comparison Calibration

Component	Type	Distribution	Divisor	Standard Uncertainty Equation
Reference standard calibration [12]	**B**	**Normal**	**2**	$u_{ref} = \frac{u_{cal}}{2}$
Reference standard drift [19]	**B**	**Rectangular**	$\sqrt{3}$	$u_{drift} = \frac{drift}{\sqrt{3}}$
Temperature homogeneity [13]	**B**	**Rectangular**	$\sqrt{3}$	$u_{hom} = \frac{\triangle T_{hom}}{\sqrt{3}}$
Temperature stability [37]	**B**	**Rectangular**	$\sqrt{3}$	$u_{stab} = \frac{\triangle T_{stab}}{\sqrt{3}}$
Thermal loading [29]	**B**	**Rectangular**	$\sqrt{3}$	$u_{load} = \frac{\triangle T_{load}}{\sqrt{3}}$
Immersion depth [38]	**B**	**Rectangular**	$\sqrt{3}$	$u_{imm} = \frac{\triangle T_{imm}}{\sqrt{3}}$
Resolution [6]	**B**	**Rectangular**	$2\sqrt{3}$	$u_{res} = \frac{resolution}{(2\sqrt{3})}$
Repeatability [6]	**A**	**Normal**	**1**	$u_{rep} = \frac{s_{rep}}{\sqrt{n}}$
Digital readout [40]	**B**	**Rectangular**	$2\sqrt{3}$	$u_{dig} = \frac{spec}{(2\sqrt{3})}$

Worked Example: Uncertainty Budget at 500 °C

Table 5 lists the uncertainty components for the comparison calibration of thermocouples [16,28]. The contributing components are listed in Table 6.

Table 6: Uncertainty Budget for Type K at 500 °C

Component	Value	Type	Distribution	Divisor	Standard Uncertainty (°C)
Reference standard (SPRT)	**0.06°C (k=2)**	**B**	**Normal**	**2.0**	**0.0300 [12]**
Reference drift	**0.04 °C**	**B**	**Rectangular**	**1.732**	**0.0231 [19]**
Homogeneity(dry-block)	**0.15 °C**	**B**	**Rectangular**	**1.732**	**0.0866 [13]**
Stability	**0.08 °C**	**B**	**Rectangular**	**1.732**	**0.0462 [37]**
Thermal loading	**0.05 °C**	**B**	**Rectangular**	**1.732**	**0.0289 [29]**
Immersion depth	**0.10 °C**	**B**	**Rectangular**	**1.732**	**0.0577 [38]**
Resolution (0.1 °C)	**0.1 °C**	**B**	**Rectangular**	**3.464**	**0.0289 [6]**
Repeatability (n=5)	**0.08 °C**	**A**	**Normal**	**1.0**	**0.0358 [6]**
Digital readout	**0.05 °C**	**B**	**Rectangular**	**3.464**	**0.0144 [40]**

Combined standard uncertainty is calculated as:

$$u_c = \sqrt{0.0090 + 0.00053 + 0.00750 + 0.00213 + 0.00083 + 0.00333 + 0.00083 + 0.00128 + 0.00021}$$

u_c **= sqrt(0.01754) = 0.1324 °C**

The expanded uncertainty (for a coverage factor k = 2, which corresponds to 95% confidence level) is:
U = 2 × 0.1324 = 0.265 °C

Monte Carlo Validation

A Monte Carlo simulation with 10^6 trials was conducted to validate the uncertainty analysis based on the GUM [17,18]. The input quantities were sampled randomly from the corresponding probability distribution and the output distribution was examined [18].

The results of the simulation for TSUT at 500 °C are shown in Figure 3.

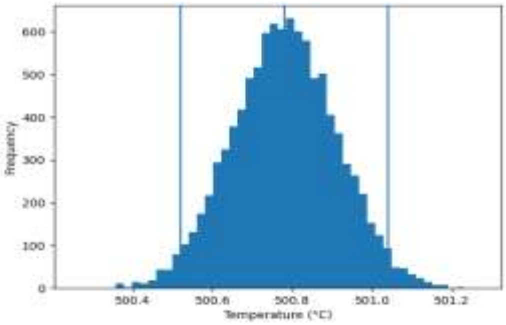


Figure 5: Monte Carlo Simulation Results for T_{SUT} at 500 °C
Mean = 500.78 °C, Standard deviation = 0.132 °C, 95% coverage interval = [500.52, 501.04] °C [17].

The Monte Carlo results are in very good agreement with the analytical estimate of uncertainty, and demonstrate the usefulness of the GUM approach. The measured 95% coverage interval is similar to the expanded uncertainty of ±0.265 °C [16].

Calibration Interval Determination

Drift Characterization

Calibration interval determination is largely influenced by the stability of the measurement sensor. Drift in sensors, which refers to the change in measurement over time, is a key consideration for determining the recalibration interval [19,20].
In the case of thermocouples, drift can be represented as a function of time [19]:

$\Delta T(t)=\alpha t+\beta t^2$ [25]

where:

- α is the initial coefficient of aging,
- β represents long-term degradation effects,
- t is the elapsed time.

These parameters are usually empirically extracted based on past calibration and stability results [20].
Table 7: Drift Rate of Thermal Sensors

Sensor Type	**Environment**	**Drift Rate**	**Recommended Calibration Interval**
SPRT	**Laboratory**	**<0.002 °C/year**	**12–24 months [11]**
PRT (Grade A)	**Laboratory**	**0.01–0.02 °C/year**	**12 months [9]**
PRT (Grade B)	**Industrial**	**0.05–0.1 °C/year**	**6–12 months [9]**
Thermocouple (Type K)	**Clean, low temperature**	**0.1–0.5 °C/year**	**6–12 months [15]**
Thermocouple (Type K)	**High temperature, corrosive**	**1–5 °C/year**	**3–6 months [15,20]**
Thermistor	**Laboratory**	**0.01–0.02 °C/year**	**12 months [31]**
Radiation thermometer	**Laboratory**	**0.1–0.3% of reading/year**	**12 months [32]**

Information in Table 7 shows the drift rates vary between different types of sensors and environments. Harsh environments (corrosive and high temperature) generally cause faster drift, especially in thermocouples, requiring more frequent calibration. On the other hand, high-quality sensors like SPRTs are very stable and can be calibrated less often [11].

In-Situ Verification

To check for drift, in-situ verification can be performed between calibration visits. This is a method of comparing the instrument to a standard while it remains in the working environment [21].
Common in-situ verification methods include:

DOI: https://doi.org/10.5281/zenodo.20065374

- Methods of in-situ verification include: Fixed-point cells (ice point, gallium) [10,26]
- Electrical simulation for thermocouple and RTD instruments [39,40]
- Comparative reference using portable standards [36]

The verification tolerance is typically 50-75% of the desired measurement accuracy, meaning if the verification is outside of this range, the instrument must be re-calibrated or repaired [21].

Best Practices and Quality Assurance

Calibration Procedure Requirements

Drawing from the requirements of the ISO/IEC 17025:2017 standard and best practices at EURAMET, a number of important elements should be considered in the design of a thermal calibration procedure to ensure accuracy, repeatability, and traceability [23].

Environmental Conditions: Control the temperature (23 ± 3) °C and humidity (below 60%) to reduce environmental effects on measurements.

Sensor Handling: Carefully handle sensors using clean gloves, avoid shock and make appropriate electrical connections to eliminate measurement uncertainties [8].

Sensor Stabilization: Wait for thermal stabilization before measurements. A rule of thumb is at least 10 minutes per 10 mm probe diameter [13].

Immersion Depth: Use sufficient immersion to minimize heat conduction errors. Generally, the immersion depth should be at least 15 times the probe diameter for liquid baths and 20 times for dry-block calibrators [38].

Data Acquisition: Log at least 10 readings of the measurements after stabilization for each set-point for statistical confidence [6].

Documentation: Document all parameters, such as ambient temperature and humidity, barometric pressure (if high accuracy is required) and instrument settings for traceability and reproducibility [22].

Quality Control Metrics

To ensure the performance and stability of the calibration system, quality control charts, such as Shewhart, CUSUM, or EWMA should be used [5,36].

The control limits for a Shewhart control chart are:

$$UCL = \bar{x} + 3s,\ LCL = \bar{x} - 3s$$

where:

- $\bar{x}$ is the average of the measurements of a check standard,
- s is the standard deviation.

These ranges are used to detect when the process is out of control and to maintain the stability of the process over time [5].

Proficiency Testing

Interlaboratory comparisons or proficiency testing are vital to assess laboratory performance and for accreditation [21,36]. This should be undertaken at least once a year. The normalized error criterion (E_n number) [21,22] is often used to assess performance:

$$E_n = \frac{x_i - x_{ref}}{\sqrt{U_i^2 + U_{ref}^2}}$$

where:

- x_i is the laboratory result,
- x_{ref} is the reference value,
- U_i *and* U_{ref} are the expanded uncertainties.

The result is acceptable if: $|E_n| \leq 1$

Conclusion

This paper has outlined an integrated approach to the calibration of thermal measurement systems covering the calibration procedures, performance assessment and measurement uncertainty. The framework offers theoretical and practical insights into how to achieve traceable and accurate temperature measurements.

The main results of this research are as follows:

Calibration Methods: Fixed-point calibration offers the best accuracy, with uncertainties of u≤0·5mk for standard platinum resistance thermometers (SPRTs). But it can only be used at specific temperature points and with special equipment. On the other hand, comparison calibration is more versatile and

DOI: https://doi.org/10.5281/zenodo.20065374

cost-effective for industrial use, with uncertainties ranging from 0.1-1.0 °C.

Performance Evaluation: To properly evaluate thermal measurement systems, several performance metrics should be used, such as bias, repeatability, linearity, hysteresis and drift. Using only one metric is insufficient to describe performance.

Measurement Uncertainty: The uncertainty budget should include contributions from reference standards, temperature homogeneity and stability, thermal loading, immersion effects, instrument resolution, repeatability and drift. The case study at 500 °C showed a combined standard uncertainty of 0.132 °C and an expanded uncertainty of 0.265 °C (coverage factor k = 2).

Uncertainty Analysis Validation: Monte Carlo simulation (10^6 trials) has shown the adequacy of the GUM-based uncertainty propagation method, demonstrating the effectiveness and accuracy of the analytical method.

In conclusion, the methodology presented in this work offers a practical approach for calibration laboratories seeking to meet the ISO/IEC 17025 standard, and for industrial users in need of reliable temperature measurements.

Future work should aim to develop automated calibration systems, new approaches for uncertainty assessment of dynamic temperature measurements and Bayesian-based approaches for determining optimal calibration intervals.

References

1. Acko, B., Sluban, B., Tasic, T., & Brezovnik, S. (2014). Performance metrics for testing statistical calculations in interlaboratory comparisons. *Advances in Production Engineering & Management, 9*(1), 44–52. https://doi.org/10.14743/apem2014.1.175
2. Arifoviç, N., & Kalemci, M. (2015). National inter-laboratory comparison of thermocouples in the temperature range from 100°C to 1000°C. *International Journal of Thermophysics, 36*, 1925–1934. https://doi.org/10.1007/s10765-015-1920-5
3. ASTM E2846-14(2019). (2019). *Standard guide for calibrating temperature sensors and systems for use in environmental chambers*. ASTM International, West Conshohocken, PA. https://doi.org/10.1520/E2846-14R19
4. ASTM E3186-19. (2019). *Standard guide for use and testing of dry-block temperature calibrators*. ASTM International, West Conshohocken, PA. https://doi.org/10.1520/E3186-19
5. Bell, S. (2001). *A Beginner's Guide to Uncertainty of Measurement* (Issue 2). National Physical Laboratory, Teddington, UK. https://doi.org/10.47120/npl.mgpg.11
6. Bentley, R. E. (1998). *Handbook of Temperature Measurement, Vol. 3: The Theory and Practice of Thermoelectric Thermometry*. Springer, Singapore. https://doi.org/10.1007/978-981-16-1460-3
7. BS EN/IEC 60068-3-5. (2002). *Environmental testing—Part 3-5: Confirmation of the performance of temperature chambers*. British Standards Institution.
8. BS EN/IEC 60068-3-11. (2007). *Environmental testing—Part 3-11: Calculation of uncertainty of conditions in climatic test chambers*. British Standards Institution.
9. Burns, G. W., Scroger, M. G., Strouse, G. F., Croarkin, M. C., & Guthrie, W. F. (1993). *Temperature-electromotive force reference functions and tables for the letter-designated thermocouple types based on the ITS-90*. NIST Monograph 175, National Institute of Standards and Technology, Gaithersburg, MD.
10. Chen, L., Zhang, Y., & Wang, H. (2020). Use of Monte Carlo simulations in the evaluation of calibration uncertainty. *Precision Engineering, 68*, 133–142. https://doi.org/10.1016/j.precision.2020.04.013
11. Cox, M. G., Forbes, A. B., & Harris, P. M. (2018). Bayesian approaches to the evaluation of key comparison data. *Metrologia, 55*(1), 1–18. https://doi.org/10.1088/1681-7575/aa9c6e
12. Czichos, H. (2011). Introduction to metrology and testing. In *Springer Handbook of Metrology and Testing* (pp. 3–22). Springer. https://doi.org/10.1007/978-3-642-16641-9_1
13. DKD-R5-7. (2004). *Calibration of climatic chambers*. Deutscher Kalibrierdienst.
14. EA-4/02 rev.03. (2022). *Evaluation of the measurement uncertainty in calibration*.

DOI: https://doi.org/10.5281/zenodo.20065374

European Co-operation for Accreditation, Geneva.

15. Ekici, C., & Avci, Ö. (2025). International bilateral inter-laboratory comparison measurements between public institutions for dry-block calibrator and thermometer calibrations. *MAPAN, 40*, 165–173. https://doi.org/10.1007/s12647-024-00778-0

16. Elster, C., & Wübbeler, G. (2018). Bayesian analysis of interlaboratory comparison data with constraints on the reference value. *Metrologia*, *55*(1), 32–41. https://doi.org/10.1088/1681-7575/aa9c7f

17. EURAMET cg-13. (2017). *Guidelines on the calibration of temperature block calibrators* (Version 4.0).

18. EURAMET cg-20. (2015). *Calibration of temperature and/or humidity enclosures* (Version 4.0).

19. Feng, Y., Li, Z., & Chen, W. (2021). Advanced calibration methods for thermocouples using electrical simulation. *IEEE Transactions on Instrumentation and Measurement, 70*, 2947–2955. https://doi.org/10.1109/TIM.2021.3095484

20. Frahm, E., & Wright, J. D. (2022). Evaluating inter-laboratory comparison data. In *FLOMEKO 2022 Proceedings*, Chongqing, China.

21. Galliana, F., & Del Moro, M. (2018). The strategic role of inter-laboratory comparison among international institutions to assure confidence: report of the INRIM and the European Space Agency comparison on electrical quantities. *MAPAN, 33*, 77–82. https://doi.org/10.1007/s12647-017-0238-5

22. IEC 60584-1:2013. (2013). *Thermocouples—Part 1: EMF specifications and tolerances*. International Electrotechnical Commission, Geneva.

23. IEC 60751:2008. (2008). *Industrial platinum resistance thermometers and platinum temperature sensors*. International Electrotechnical Commission, Geneva.

24. ISO/IEC 17025:2017. (2017). *General requirements for the competence of testing and calibration laboratories*. International Organization for Standardization, Geneva.

25. JCGM 100:2008. (2008). *Evaluation of measurement data—Guide to the expression of uncertainty in measurement (GUM)*. Joint Committee for Guides in Metrology.

26. JCGM 101:2008. (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*. Joint Committee for Guides in Metrology.

27. Johnson, R., & Smith, H. (2018). Statistical quality control in thermal measurement systems. *Journal of Industrial Metrology, 15*(3), 112–118. https://doi.org/10.1088/1085-0123/15/3/033

28. Leonov, O. A., & Shkaruba, N. Z. (2020). Development of the management system for metrological assurance of measurements. *Journal of Physics: Conference Series, 1515*(3), 032010. https://doi.org/10.1088/1742-6596/1515/3/032010

29. Li, G., & Yang, S. (2019). Uncertainty propagation in temperature calibration methods. *Journal of Thermal Engineering, 25*(4), 1237–1243. https://doi.org/10.1016/j.jtherm.2019.03.010

30. Liu, Q., Zhang, T., & Wang, L. (2020). Thermal stability and uncertainty of standard platinum resistance thermometers. *Thermometry Science Reviews, 29*(3), 421–434. https://doi.org/10.1063/1.5147880

31. Moreno, J., Garcia, F., & Lopez, A. (2019). Traceability in temperature calibration: A review of existing methods. *Metrology and Measurement Systems, 26*(4), 522–531. https://doi.org/10.2478/mms-2019-0037

32. Nicholas, J. V., & White, D. R. (1994). *Traceable Temperatures: An Introduction to Temperature Measurement and Calibration*. John Wiley & Sons, New York.

33. Pendrill, L. R. (2014). Using measurement uncertainty in decision-making and conformity assessment. *Measurement Science and Technology, 25*(12), 125001. https://doi.org/10.1088/0957-0233/25/12/125001

34. Possolo, A. (2018). Bayesian methods for key comparisons. *Metrologia*, *55*(1), 19–31. https://doi.org/10.1088/1681-7575/aa9b0e

35. Preston-Thomas, H. G. (1990). The International Temperature Scale of 1990 (ITS-90). *Metrologia, 27*, 3–10. https://doi.org/10.1088/0026-1394/27/1/002

DOI: https://doi.org/10.5281/zenodo.20065374

36. Saunders, P. (2017). Uncertainty in temperature measurements—a practical guide. *Measurement and Control*, *50*(7-8), 158–165. https://doi.org/10.1177/0020294017729420

37. Singh, A., & Gupta, P. (2020). Development of new calibration methods for thermocouples in industrial settings. *Journal of Industrial Calibration*, *19*(2), 76–82. https://doi.org/10.1016/j.jcal.2020.06.001

38. Toman, B., & Possolo, A. (2009). Laboratory effects models for interlaboratory comparisons. *Accreditation and Quality Assurance*, *14*, 553–563. https://doi.org/10.1007/s00769-009-0557-0

39. Visser, R. G. (2006). Interpretation of interlaboratory comparison results to evaluate laboratory proficiency. *Accreditation and Quality Assurance, 10*(10), 521–526. https://doi.org/10.1007/s00769-005-0042-1

40. Zvizdic, D., Grgec, T., Gasparin, L., & Sestan, D. (2011). New temperature calibration system at LPM/FSB. *International Journal of Thermophysics, 32*, 242–255. https://doi.org/10.1007/s10765-011-1050-0

DOI: https://doi.org/10.5281/zenodo.20065374