Journal of the Chinese Society of Mechanical Engineers, Vol.47, No.4, pp319~328 (2026)

A Directional Consistency Analysis Framework for Performance Assessment of CMMs

Jr-Rung Chen*, Jui-Hung Liu*, Sheng-Hsiang Hung** and Ming-Xian Lin***

Keywords : coordinate measuring machine, measurement system analysis, CMM monitoring, directional coverage ratio.

ABSTRACT

Leading manufacturers require consistent measurement results between their own coordinate measuring machines (CMMs) and those of their suppliers, but calibration certificates and aggregate Gage R&R provide limited direction-specific diagnostic information. This study proposes a directional consistency analysis framework combining calibrated-reference deviations, leave-one-out peer comparison, directional performance indicators, and cross-feature comparison using the Directional Coverage Ratio (DCR) and Jaccard index. Five CMMs measured a 400 mm gauge block and a 30 mm reference sphere three times in seven directions. For the monitored supplier CMM, the directional error span was 12.23 μm, the exceedance ratio was 42.9%, and the normalized directional span was 2.23. Repeat-based 95% intervals identified Y, ZX+45°, and ZX−45° as the main alarm directions outside the ±2.743 μm screening band. The sphere-form alarm set covered all three gauge-block alarms, resulting in DCR = 1.00 and a Jaccard index of 0.429. A sensitivity analysis over screening bands from ±2.0 to ±3.5 μm showed that the maintenance-priority classification remained unchanged. The framework can therefore serve as a complementary tool for routine CMM monitoring, including health status and supplier measurement consistency assessments in addition to ISO-based verification, formal Gage R&R, and measurement uncertainty evaluation.

Paper Received July, 2026. Revised August, 2026. Accepted August, 2026. Author for Correspondence: Ming-Xian Lin.

* *Department of Intelligent Automation Engineering, National Chin-Yi University of Technology, Taichung, Taiwan, ROC.*

** *Precision Mechanical Metrology Research Laboratory, Center for Measurement Standards, Industrial Technology Research Institute, Hsinchu, Taiwan, ROC.*

*** *Department of Mechanical and Computer-Aided Engineering, Feng Chia University, Taichung, Taiwan, ROC.*

INTRODUCTION

Coordinate Measuring Machines (CMMs) are widely used in modern manufacturing because they provide high-precision measurements of dimensional and geometric characteristics, such as diameter, angle, flatness, and position (Chen et al., 2025; Lin and Hsieh, 2023; Hsieh et al., 2025; Hsu et al., 2023). As manufacturing tolerances become increasingly stringent, the reliability and traceability of CMM measurements have become critical to product quality, making performance verification and periodic calibration essential. To ensure measurement accuracy and reliability, international standards such as ISO 10360 and ISO 15530 provide standardized procedures for CMM performance verification and uncertainty evaluation (Štrbac et al., 2017; Petrò and Moroni, 2021). In particular, ISO 10360-2 specifies multi-position and multi-length measurement tests to assess CMM performance throughout the measuring volume. Although laser interferometer-based calibration methods can achieve extremely high accuracy, their complexity and cost often limit industrial adoption. Consequently, calibrated reference standards such as gauge blocks, sphere plates, hole plates, and ring gauges remain practical and cost-effective solutions for routine verification, calibration, and measurement traceability in industrial applications.

Inter-laboratory comparisons are an important approach for verifying CMM measurement capability and ensuring the accuracy, consistency, and traceability of measurement results (Hansen and De Chiffre, 1999; Arenhart et al., 2021). Through these comparisons, laboratories can validate measurement methods, evaluate measurement uncertainty, and identify potential sources of discrepancy. The International Laboratory Accreditation Cooperation (ILAC) provides a framework for mutual recognition, ensuring that participating laboratories comply with internationally accepted measurement standards. These proficiency tests not only demonstrate measurement competence but also support the continuous improvement of technical capabilities and measurement reliability. To ensure meaningful comparison results, standardized measurement procedures and detailed documentation are required, including measurement locations, orientations, instruments, reference standards, and measurement ranges (Gawor et al., 2021; Buchta et al., 2025). In addition, factors such as visual inspection, cleaning procedures, temperature, humidity, thermal stabilization time, and fixture setup must be properly controlled. Comprehensive records of these conditions provide an essential basis for uncertainty evaluation, traceability, and the identification of necessary corrective actions.

Measurement uncertainty is a key factor in ensuring the accuracy and reliability of CMM measurements. Previous studies have investigated uncertainty sources and evaluation methods, including instrument calibration, environmental conditions, and probing system performance (Hsieh et al., 2024). Bell (2001) identified measuring instruments, workpieces, environmental influences, and operators as major contributors to measurement uncertainty and emphasized the importance of traceable calibration and proper uncertainty evaluation. Takatsuji et al. (2002) developed a CMM calibration model based on ISO GUM and showed that probe uncertainty is a dominant contributor to measurement error, whereas environmental effects are relatively minor. Jakubiec et al. (2012) proposed an uncertainty estimation method considering geometric errors, probe inaccuracies, and temperature variations in accordance with ISO GUM guidelines. In addition to uncertainty evaluation, CMM performance verification is essential for maintaining measurement accuracy. Rahman and Mayer (2016) categorized verification methods into mechanical artefacts, optical techniques, and opto-mechanical approaches, with calibrated reference standards remaining the most widely adopted solution in industry. Fesperman et al. (2010) further demonstrated that machine positioning errors can significantly affect probing performance, highlighting the importance of proper calibration. More recently, Sato et al. (2024) proposed a practical experimental method for estimating uncertainty in CMM-based dimensional and form measurements. Their method uses analysis of variance to formulate measurement uncertainty and was validated through comparison experiments, demonstrating consistency with established calibration approaches. Building on these studies, ISO/TS 15530-1 specifies uncertainty evaluation methods based on calibrated reference standards, simulation, and sensitivity analysis, while ISO 15530-3 provides an experimental approach that is widely used in industrial practice. Overall, calibrated reference standards and tactile CMMs remain important tools for high-accuracy measurement, performance verification, and uncertainty evaluation in industrial applications.

Gage Repeatability and Reproducibility (Gage R&R) is widely used to check the precision of a measurement system. It estimates repeatability, reproducibility, and part-to-part variation through crossed or nested study designs. In industry, the estimated gage variation, its percentage of total study variation or tolerance, and the number of distinct categories are commonly reported to judge whether a measurement system is suitable for process control or product acceptance (Vardeman and VanValkenburg, 1999; Burdick et al., 2003). However, aggregate Gage R&R results provide limited diagnostic information. They may not reveal specific error sources such as geometric error, compensation deficiency, probe drift, or feature-dependent measurement behavior unless these factors are deliberately included in the study design. In addition, a CMM may have good repeatability while still producing biased results relative to a calibrated reference standard. More detailed spatial studies can improve diagnostic capability, but they require many repetitions and controlled randomization. This makes them difficult to apply during routine supplier monitoring.

This study proposes a directional consistency framework for CMM performance checks using calibrated reference standards and peer comparison. Unlike ISO-based verification and uncertainty evaluation, which primarily address measurement performance and traceability, and Gage R&R, which focuses on measurement-system variation, the proposed framework provides direction-specific diagnostic information. Conventional inter-machine comparisons can identify differences among CMMs but generally provide limited guidance regarding which measurement directions or system components should be inspected. The proposed framework integrates calibrated-reference deviations, leave-one-out peer comparison, directional variation, measurement repeatability, and cross-feature alarm comparison. This combination enables abnormal measurement behavior to be traced to specific directions and converted into practical inspection priorities. Therefore, the framework is intended as a complementary diagnostic tool for routine supplier CMM monitoring rather than a replacement for existing CMM performance assessment methods.

METHODS

A schematic structure of the diagnostic framework is presented in Figure 1. The framework consists of two complementary layers: a reference layer based on calibrated reference standards and a peer layer composed of different CMMs. The reference layer determines whether a measurement exceeds the screening band established for the comparison, whereas the peer layer evaluates whether the target CMM exhibits an outlier response relative to the other machines in the same measurement direction. A cross-feature check is subsequently performed to determine whether the alarm directions identified by the reference sphere are consistent with those detected in the length measurements. In this study, five CMMs participate in the comparison, including three CMMs from the lead manufacturer (S-ADC, S-VAST XT, and S-RDS), one monitored supplier CMM (KH), and one advisory-laboratory CMM (ITRI). The lead-manufacturer and supplier systems are anonymized to avoid disclosing manufacturer and model information, while ITRI serves as an independent technical peer. The role of each participant is summarized in Table 1.

The two diagnostic layers provide complementary evidence for system evaluation. A measurement outside the reference screening band together with a peer alarm represents the highest-confidence alarm state, indicating that the target machine both exceeds the reference criterion and behaves abnormally relative to the peer group. A reference-layer alarm without a corresponding peer alarm remains significant because the peer distribution may be broad or shifted. Conversely, a peer alarm within the reference screening band suggests an emerging abnormal trend that warrants protocol review rather than an immediate declaration of nonconformity. Each measurement direction is therefore classified into one of these four evidence states before system-level indicators are evaluated. Instead of producing a single weighted health score, the framework reports direction-level evidence, including the number and identity of out-of-band directions, the magnitude and sign of peer-deviation indices, the error range across all directions, and the consistency of alarms across different measurement features. This approach preserves traceability to individual measurement directions while allowing the lead manufacturer to refine escalation rules without recalculating the original measurement data.

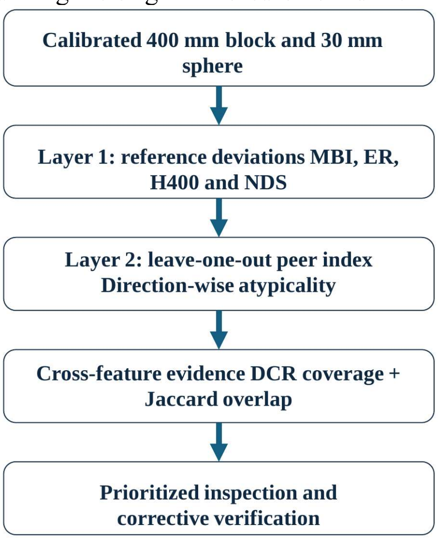

Fig. 1. Schematic structure of two-layer diagnostic monitoring.

Table 1. Roles in the industrial monitoring case.

Role	Case unit	Diagnostic purpose
Lead manufacturer	S-ADC, S-VAST XT, S-RDS	Defines monitoring need and internal peer data.
Supplier	KH	Target CMM; excluded from its own peer deviation index.
Advisory laboratory	ITRI	Independent technical peer.
Reference basis	Calibrated reference standards	Assigned values for reference deviations.

The analysis was based on three repeated measurements for each CMM system, measured feature, and spatial direction. Since laboratory-grade CMMs are usually high-quality machines, three repetitions are sufficient to capture short-term repeatability and help detect directional anomalies. This method is consistent with the practical testing concept of ISO 10360-2 and also considers the need for efficiency during on-site machine verification. This study utilized two calibrated reference standards: a 400 mm gauge block and a 30 mm reference sphere (Figure 2). The 400 mm gauge block was measured using the point-to-plane method. Four points were measured on one end face to define a reference plane. One point was then measured at the center of the opposite end face. The normal distance from this point to the reference plane was calculated. Measurements were made in seven directions. These directions were X, Y, Z, ZX, ZY, ZX+45°, and ZX−45°. The ZX+45° and ZX−45° directions were diagonal directions in the ZX plane. The reference sphere was measured by least-squares sphere fitting. The measurement system was programmed to automatically probe each point on the gauge block and reference sphere in a specified sequence, with designated repetitions, calculating corrections as needed (Šafarič et al., 2021). The same direction labels were used for the gauge block and the reference sphere. The CMM systems used by S-ADC, S-VAST XT, S-RDS, KH, and ITRI all had temperature measurement and compensation. Each participant used its own CMM temperature-compensation setting. The reported results were used directly. No additional temperature correction was applied in this study. The same direction labels were used to support direction-based comparisons, making it possible to compare results across different CMM systems and artefacts. The measurement setup reflects a routine laboratory CMM check. Each participant used one calibrated reference standard for each measured feature. Each participant also followed its own standard operating procedure and operator arrangement. The data therefore show the measurement behavior under routine check conditions. The data were used for inter-system comparison, direction-based screening, and short-term repeatability assessment.

The assessment of measurement uncertainty for the CMM check method follows the guidelines provided in the ISO Guide to the Expression of Uncertainty in Measurement (GUM) (Lin and Hsieh, 2023; ISO/IEC Guide 98-3, 2008). The three repeated measurements were used to describe short-term variation. They were also used to calculate the within-condition standard deviation. A Student's t-based confidence interval for the mean was also calculated. The participant-supplied measurement uncertainty budgets were not used in the direction-screening indicators. Therefore, the t-based confidence intervals in this study describe only short-term repeated measurements. They should not be treated as complete measurement uncertainty intervals.

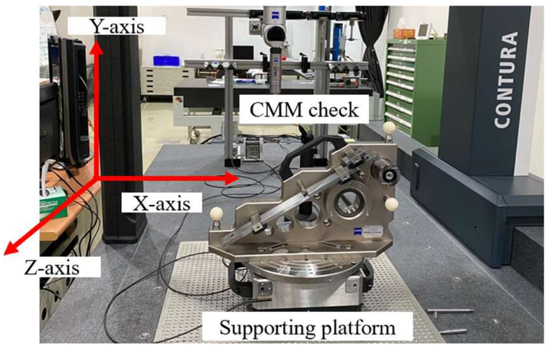

Fig. 2. Artefacts of gauge block and measuring set-up.

For each measurement parameter, this study collected 105 readings — five CMM systems × seven directions × three repeats — which were reduced to 35 direction-averaged values per parameter (five systems × seven directions), each obtained from its three repeated readings. The three parameters were the length of the 400 mm gauge block and the diameter and form of the reference sphere; the measured feature, reported measurand, and primary diagnostic use of each artefact are summarized in Table 2. This structure supports comparison of inter-machine errors across the seven directions, together with short-term repeatability assessment and peer comparison. Because all participants used the same artefacts and procedures, shared environmental and part-related conditions may influence the results and should be taken into account when interpreting them.

Table 2. Artefacts and primary diagnostic use.

Feature	Reported measurand	Primary use
400 mm gauge block	Point-to-plane distance	Long-length and direction response.
30 mm reference sphere	LSQ diameter and form	Secondary cross-feature screening.

For machine m and direction d, the reference deviation equals the reported value minus the calibrated reference value:

$$e_{m,d} = x_{m,d} - r_d. \tag{1}$$

The campaign used the same two-sided screening band in every direction, L_{400} = 2.743 µm. This value was an existing screening criterion in the original industrial comparison dataset. However, the available analysis records did not document its derivation or establish a direct relationship with participant-supplied calibration uncertainty, a coverage factor, maximum permissible error, guard-banding criteria, or a specific ISO requirement. Therefore, L_{400} is treated in this study as an operational screening band rather than a formal metrological conformity limit. To evaluate the influence of this threshold on the diagnostic results, a sensitivity analysis was performed using screening bands from ±2.0 to ±3.5 µm to determine whether the identified alarm directions, machine ranking, and maintenance-priority classification remained consistent. This study uses three indicators to describe the performance of each machine across the seven directions. The high-risk set $H_{400,m}$ contains directions where $|e_{m,d}| > L_{400}$. MBI gives the mean absolute normalized deviation, and ER gives the fraction of directions outside the screening band:

$$\mathrm{MBI}_m = \frac{1}{D}\sum_{d=1}^{D}\frac{|e_{m,d}|}{L_{400}}, \qquad \mathrm{ER}_m = \frac{|\{d : |e_{m,d}| > L_{400}\}|}{D} \tag{2}$$

NDS describes the range of deviations across different directions. It is defined as the difference between the maximum and minimum deviations among all directions, divided by the total width of the two-sided screening band:

$$\mathrm{NDS}_m = \frac{[\max_d(e_{m,d}) - \min_d(e_{m,d})]}{2L_{400}} \tag{3}$$

From a diagnostic point of view, a high MBI means that the average error is large. A high ER means that the measurement problem appears in many spatial directions, rather than only in one direction. A high NDS means that the errors are strongly different among directions. This type of error pattern may be related to geometric errors in the CMM structure, such as straightness errors, squareness errors, or other orientation errors. These errors can cause direction-dependent measurement deviations (Schwenke et al., 2008).

MBI, ER, and NDS describe complementary aspects of measurement performance. MBI represents the average deviation, ER represents the proportion of out-of-band directions, and NDS represents the directional error range. NDS is insensitive to a common shift because it is based on the error range. If all directional errors remain within ±L400, NDS cannot exceed 1. Values greater than 1 indicate that the observed error range is wider than the screening band. Unlike MBI and ER, NDS does not identify the responsible direction. For each target CMM and measurement direction, a leave-one-out strategy was applied by excluding the target machine from the peer statistics. The peer mean and sample standard deviation were calculated using the remaining four CMMs, providing an independent baseline for comparison. The LOO peer-deviation indicator is defined as follows.

$$p_{m,d} = \frac{e_{m,d} - \bar{e}_{-m,d}}{s_{-m,d}} \tag{4}$$

where $e_{m,d}$ is the deviation of machine m in direction d. The term $\bar{e}_{-m,d}$ is the average deviation of the other peer machines in direction d. The term $s_{-m,d}$ is the sample standard deviation of those other peer machines in direction d. This study uses an absolute value of 2 as the operational screening limit. Each comparison involves only four peer machines. Also, we did not use the uncertainty budgets from participants for weighted calculations. The LOO indicator therefore mainly reflects how far the target machine deviates from the other CMMs in a given direction. A small sample standard deviation among peer machines might amplify the LOO indicator. Accordingly, it should not be treated as a standard statistic under a normal distribution. It serves only to screen for relative directional deviations. The leave-one-out method ensures the target machine does not affect its own baseline. As a result, every alarm traces back to clear directional deviation data.

The sphere form alarm set $H_{\text{sphere,m}}$ is the group of directions where machine m exceeds the limit for sphere form. This study uses a limit of about 1.600 µm for sphere form. The reference values for sphere form change slightly depending on the direction. The values are 0.100 µm for the X through ZY directions. The value is 0.120 µm for the ZX+45° direction. The value is 0.130 µm for the ZX−45° direction. We kept these exact original values when we organized the data. We used a two-sided limit of ±1.685 µm for the sphere diameter. We used these limits to find alarm directions and check for directional matches. We calculated the DCR using the sphere form alarm set. We used the sphere diameter results as extra information. If a machine has at least one high-risk direction for the 400 mm block, its DCR is:

$$\text{DCR}_m = \frac{|H_{\text{400,m}} \cap H_{\text{sphere,m}}|}{|H_{\text{400,m}}|} \tag{5}$$

where $H_{\text{400,m}}$ is the set of high-risk directions for machine m when measuring the 400 mm block. $H_{\text{sphere,m}}$ is the set of alarm directions for the same machine when measuring the sphere form. We do not calculate the DCR if $H_{\text{400,m}}$ is empty. The DCR shows a specific ratio. It quantifies how many of the 400 mm high-risk directions also appear in the sphere form alarm set. Thus, the DCR helps us see if the long-length errors and sphere form errors happen in the same directions. A single ratio might be misleading. Therefore, we also report the number of directions in $|H_{\text{400,m}}|$ and $|H_{\text{sphere,m}}|$. We also report the number of overlapping directions and the Jaccard index. Sometimes $H_{\text{sphere, m}}$ covers all directions. If this happens and $H_{\text{400,m}}$ is not empty, the DCR will be 1. In this case, we must look at the DCR alongside the 400 mm error profile and the Jaccard index. All calculations use the original numbers without rounding.

Before setting inspection priorities, this study examined the interval information from the three repeated measurements. For each KH feature-direction combination, we calculated a two-sided 95% Student's t confidence interval for the short-term mean. The formula is $\bar{x} \pm t_{0.975,2}\, s/\sqrt{3}$. Here, $t_{0.975,2} = 4.303$, and s is the sample standard deviation of the three repeated measurements. We calculated this interval directly from the three repeated readings. It describes the short-term variation of the mean under the same measurement conditions. In other words, this interval mainly shows the short-term repeatability. It serves as extra information for directional screening and sorting inspection priorities. The original measurement data, temperature compensation methods, and directional consistency were first verified. The reference-layer indicators, leave-one-out peer comparison, and directional coverage indicators were then calculated. Finally, evidence from the reference layer, peer layer, and repeatability analysis was integrated to establish inspection priorities. Alarms concentrated in specific axial or diagonal directions were prioritized for inspection, whereas widespread sphere alarms prompted further verification of the probe system and sphere-fitting settings. The resulting inspection priority list provides practical guidance for measurement confirmation and subsequent system checks.

RESULTS AND DISCUSSION

As shown in Figure 3 and Table 3, the 400 mm gauge block results indicate that KH should be the first priority CMM for inspection. Among the five systems, KH exhibits the largest directional variation, with a deviation range of 12.23 µm across the seven measurement directions. The mean deviations are 1.28, −5.88, −2.38, −0.21, −2.01, 6.35, and −4.48 µm for the X, Y, Z, ZX, ZY, ZX+45°, and ZX−45° directions, respectively. The largest positive deviation occurs in the ZX+45° direction, whereas the largest negative deviation occurs in the Y direction. Based on the ±2.743 μm screening limit, the high-risk direction set for KH is $H_{400,KH} = \{Y, ZX+45^{\circ}, ZX-45^{\circ}\}$. By comparison, S-VAST XT exceeds the screening limit only in the Y direction, S-RDS exceeds the limit in the ZX and ZY directions, whereas all seven directions for S-ADC and ITRI remain within the screening band. Table 3 further shows that KH has the largest MBI (1.178), ER (0.429), and NDS (2.230), confirming that it should receive the highest inspection priority. Although the signed mean deviation of KH is only −1.050 μm, its directional standard deviation reaches 4.057 μm, indicating substantial directional inconsistency. In contrast, ITRI has a mean deviation of −1.449 μm but a deviation range of only 1.114 μm, a directional standard deviation of 0.379 μm, and an NDS of 0.203. These results indicate that ITRI exhibits a nearly uniform negative shift across all directions, whereas KH simultaneously shows large positive and negative deviations, suggesting that its primary problem is directional inconsistency rather than a common systematic shift.

The repeated measurement results further support this diagnosis. For KH, the sample standard deviations of the three repeated measurements range from 0.153 to 1.400 μm across the seven directions, whereas all repeated measurement standard deviations for the other four CMMs remain below 0.208 μm. The 95% Student's t confidence intervals for KH in the Y, ZX+45°, and ZX−45° directions are [-7.057, −4.709], [2.872, 9.828], and [-5.108, −3.858] μm, respectively, all of which lie completely outside the ±2.743 μm screening band. In contrast, the confidence interval for the Z direction is [-5.587, 0.820] μm, which crosses the lower screening limit and is therefore identified as a secondary confirmation item. Furthermore, all three individual measurements in the Y, ZX+45°, and ZX−45° directions exceed the screening limit, indicating that the identified alarm directions are not caused by isolated outliers but represent consistent directional deviations.

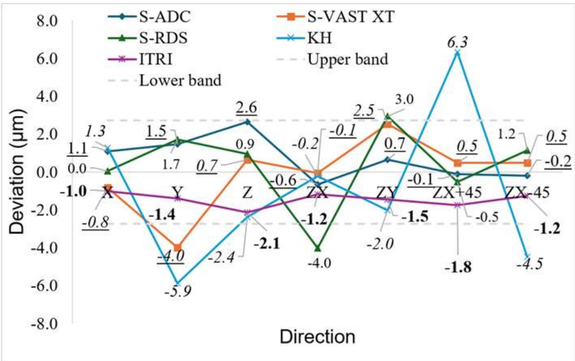

Fig. 3. Direction-wise 400 mm gauge-block deviations and screening band.

Table 3. Reference-layer indicators for the 400 mm gauge block.

System	MBI	ER	Span, μm	NDS
S-ADC	0.355	0.000	3.30	0.60
S-VAST XT	0.470	0.143	6.53	1.19
S-RDS	0.591	0.286	6.97	1.27
KH	1.178	0.429	12.23	2.23
ITRI	0.528	0.000	1.11	0.20

To assess the influence of the screening band L400 on the diagnostic conclusions, the screening band was varied across ±2.0, ±2.743, ±3.0, and ±3.5 μm and the indicators were recomputed. At ±2.0 μm, the exceedance ratios were 14.3% for S-ADC, 28.6% for S-VAST XT, 28.6% for S-RDS, 71.4% for KH, and 14.3% for ITRI. At ±2.743 μm, the corresponding values were 0.0%, 14.3%, 28.6%, 42.9%, and 0.0%. At both ±3.0 and ±3.5 μm, they were 0.0%, 14.3%, 14.3%, 42.9%, and 0.0%, respectively. The core KH alarm directions, Y, ZX+45°, and ZX-45°, remained outside every tested screening band. At ±2.0 μm, Z and ZY were additionally flagged. When the complete reference, peer, directional-span, and repeatability evidence was considered, the qualitative maintenance-priority classification remained unchanged.

As shown in Figure 4, the LOO peer comparison provides another viewpoint. This method compares each CMM with the other systems in the same direction. KH shows the most obvious relative deviation in the ZX+45° and ZX-45° directions. This matches the average deviation results of the 400 mm gauge block. At the same time, the LOO peer comparison points out that S-RDS has a local deviation in the ZX direction. This means this method can evaluate supplier CMMs and also help find local measurement differences in internal CMMs. Figure 4 shows the KH leave-one-out peer deviation indices calculated from the checked three-repeat means. The two diagonal directions were unusual. The index for ZX+45° was 7.2195, reported as 7.22, and the index for ZX-45° was -4.4440, reported as -4.44. For ZX+45°, the peer mean was -0.476 μm, the peer standard deviation was 0.945 μm, and KH reported 6.350 μm. For ZX-45°, the peer mean was 0.052 μm, the peer standard deviation was 1.020 μm, and KH reported -4.483 μm. These values are operational alarms, not formal t or z statistics. The Y direction shows why the two layers are useful together. KH was outside the screening band at -5.883 μm, but its peer deviation index was only -1.982 because the other machines varied widely in Y; the peer standard deviation was 2.694 μm. A peer-only rule would not strongly rank Y, whereas the reference layer retains it as an operational alarm. In contrast, the large diagonal indices show that KH was not only outside the common band but also unusual relative to the other machines. Across the 35 target-direction LOO comparisons, only three indices exceeded the operational threshold |p| = 2: KH at ZX+45°, KH at ZX-45°, and S-RDS at ZX. The S-RDS index was -6.978, consistent with its reference-band error of -4.017 µm. No other S-RDS direction exceeded the peer threshold. KH remained the first system-level priority because it had two opposite-signed diagonal peer alarms, three reference alarms, the largest error range, greater repeat scatter, and broad sphere-form alarms. The result also shows that the method does not assume that only the supplier machine can be wrong. The evidence matrix clarifies the direction-level decisions. KH at ZX+45° and ZX-45° triggered both the reference and peer alarms, and each alarm appeared in all three repeats. KH at Y exceeded the reference screening band but did not produce a strong peer alarm because the peer spread was large. S-RDS at ZX triggered both the reference and peer alarms, while its ZY result exceeded only the reference screening band. Directions that passed both layers remained in the trend record. KH at Z was retained for controlled confirmation because its within-cell scatter was high. This separation provides more information than one overall machine average.

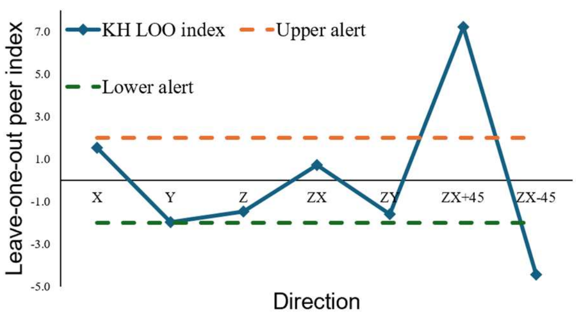

Fig. 4. KH leave-one-out peer deviation indices for the 400 mm gauge block.

As shown in Table 4 and Table 5, the DCR and Jaccard index serve as extra confirmation between different artifacts. Table 4 summarizes the form-based alarm sets. S-VAST XT and S-RDS both showed 400 mm block alarms but no sphere-form alarms, resulting in DCR = 0.00 and J = 0.000 for both systems. In contrast, the sphere-form alarms for KH cover all seven directions, resulting in DCR = 1.00. This indicates that the sphere-form alarm directions encompass all main alarm directions of the 400 mm gauge block. However, because the sphere alarm coverage is broad, this result should be evaluated alongside the Jaccard index of 0.429 and the seven-direction deviation profile of the 400 mm gauge block. S-ADC and ITRI had empty H400 sets, so DCR was not applicable. Table 5 displays the binary alarm sets in detail so that alarm states are not mistaken for magnitudes. It shows $H_{sphere,KH}$ in all seven directions and $H_{400,KH}$ only at Y, ZX+45°, and ZX-45°. The overlap therefore contains three directions. DCR = 1.00 indicates full coverage, whereas J = 0.429 indicates a broad sphere alarm that is not direction-specific. Localization comes from the 400 mm profile and repeat-based t intervals, not from DCR alone.

The source-level repeats gave KH sphere-diameter errors of -1.527, -2.460, -2.727, -2.560, -1.693, -3.060, and -2.860 µm for X, Y, Z, ZX, ZY, ZX+45°, and ZX-45°, respectively. Six of the seven values exceeded the ±1.685 µm diameter screening band. The corresponding sphere-form errors were 3.033, 2.567, 3.867, 3.633, 3.767, 4.247, and 4.537 µm. The repeat-based 95% Student's t lower bounds were 2.516, 1.942, 3.150, 3.060, 2.647, 3.622, and 3.963 µm, all above the 1.600 µm form screen. The largest non-KH form error was 1.100 µm. These intervals support the short-run repeatability of the alarm classification. However, because the screening value is undocumented, they do not support a formal conformity claim. All seven sphere-form 95% Student's t lower bounds exceeded 1.600 µm. The interval-supported H_{sphere} set therefore contained all seven directions. For KH, $H_{400,KH}$ = {Y, ZX+45°, ZX-45°}; the set results remained DCR = 3/3 = 1.00 and J = 3/7 = 0.429. Because H_{sphere} covered every direction, DCR would equal 1.00 for any nonempty H400 set. DCR therefore showed coverage but did not localize the problem. The 400 mm intervals, not DCR alone, separated the three main directions from the boundary case at Z.

Table 4. Direction-set coverage and symmetric overlap.

System	H_{400}	H_{sphere}	DCR	J
S-VAST XT	Y	None	0/1 = 0.00	0.000
S-RDS	ZX, ZY	None	0/2 = 0.00	0.000
KH	Y, ZX+45°, ZX-45°	All seven	3/3 = 1.00	0.429

Table 5. Alarm-set membership matrix for DCR and Jaccard calculations.

System	H_{400} Alarm Directions	H_{sphere} Alarm Directions	Overlap Directions
S-ADC	None	None	None
S-VAST XT	Y	None	None
S-RDS	ZX, ZY	None	None
KH	Y, ZX+45°, ZX-45°	All 7 directions	Y, ZX+45°, ZX-45°
ITRI	None	None	None

The combined evidence supports a staged inspection plan rather than a general recalibration request. Table 6 lists testable inspection candidates; they are not final causes inferred from the comparison alone. Priority was assigned without a weighted overall score. KH ranked first because it had the highest MBI, ER, error range, and NDS, together with two extreme diagonal peer indices, higher 400 mm repeat scatter, and broad sphere-form alarms supported by repeats. S-RDS ranked second for a local ZX check. It had two reference alarms and one extreme peer index but no sphere-form alarm. S-VAST XT required a narrower Y review. S-ADC and ITRI remained in routine monitoring. ITRI's relatively uniform negative pattern should still be trended even though no direction exceeded the band. This ranking converts the comparison into a clear work list and defines the measurements needed to close each action.

Overall, this study shows that the proposed framework can turn inter-system comparison data into inspection priorities. The 400 mm gauge block provides the main basis for directional findings. The three repeated measurements confirm if the alarms are stable. The LOO peer comparison determines if a CMM clearly deviates from the other systems. The DCR and Jaccard index provide extra confirmation between different artifacts. Taking KH as an example, future checks can focus first on the Y direction and the two diagonal directions of ZX+45° and ZX-45°. We should also check the probe system and sphere measurement settings at the same time.

Table 6. Evidence-based corrective priorities for KH.

Evidence	Peer index p	Inspect	Verify
ZX diagonals: 6.35 / −4.48 µm	7.22 / −4.44	Volumetric compensation; squareness; diagonal alignment; support/thermal gradient.	Repeat both diagonals; compare sign and span.
Y: −5.88 µm	−1.98	Y scale/guide; artefact alignment; local thermal gradient.	Reverse artefact; repeat with independent alignment.
Sphere form 2.57–4.54 µm (7/7); diameter screen 6/7	—	Probe qualification; stylus; reference sphere; force/filter/fit.	Requalify; repeat diameter and form map.

Note: In the Evidence and Peer index columns, paired values are listed in the order ZX+45° / ZX−45°. p is the leave-one-out peer deviation index, and $|p| > 2$ is used as the operational alarm threshold; a dash indicates that peer comparison was not applied. "(7/7)" and "6/7" denote the number of alarm directions out of the seven measured directions, based on the 1.600 µm sphere-form limit and the ±1.685 µm diameter screening band.

CONCLUSIONS

This study proposes a directional consistency analysis framework for comparing measurement results among CMM systems and establishing inspection priorities. The framework integrates calibrated-reference comparison, leave-one-out peer comparison, directional performance indicators, repeat-based assessment, and DCR/Jaccard cross-feature analysis. The experimental results show that the 400 mm gauge block can reveal long-length and direction-dependent differences among CMM systems. For KH, the deviation span across the seven directions was 12.23 µm, with Y, ZX+45°, and ZX−45° identified as the main alarm directions. The repeated measurements further supported these alarms, indicating that they were not caused by isolated measurements. The LOO peer comparison identified the largest relative deviations for KH in the ZX+45° and ZX−45° directions and also detected a local deviation for S-RDS in the ZX direction, demonstrating the capability of the framework to identify directional inconsistencies in both supplier and internal CMM systems.

The DCR and Jaccard index provide complementary cross-feature information. For KH, DCR = 1.00 indicates that the sphere-form alarms covered all main gauge-block alarm directions, whereas the Jaccard index of 0.429 indicates that the sphere-form alarm set was substantially broader. Therefore, DCR should be interpreted together with the Jaccard index and the gauge-block directional profile rather than being used independently for directional localization.

Overall, the proposed framework converts inter-system CMM comparison data into direction-specific inspection priorities. For the KH case, inspection should focus on the Y, ZX+45°, and ZX−45° directions, together with further checks of probe qualification, stylus condition, and sphere-fitting settings. The framework can serve as a practical supporting tool for routine CMM monitoring, supplier measurement consistency assessment, and targeted inspection when measurement disagreement occurs. It complements ISO-based performance verification, formal Gage R&R, and measurement uncertainty evaluation by providing additional direction-specific diagnostic information. Future work will extend the framework to additional CMM systems, reference artefacts, and measurement conditions to further evaluate its robustness and applicability in broader industrial environments.

NOMENCLATURE

D number of evaluated directions

DCR Directional Coverage Ratio

d measurement direction

e reference deviation, µm

ER exceedance ratio

H_{400} set of 400 mm gauge-block high-risk directions

H_{sphere} set of sphere-form alarm directions

L_{400} 400 mm two-sided screening band, µm

J Jaccard overlap index

LSQ least-squares

NDS normalized directional span

m CMM or measurement-system index

MBI mean absolute normalized deviation

p leave-one-out peer deviation index

r calibrated reference value, µm

s leave-one-out peer sample standard deviation, µm

x reported measurement value, µm

ACKNOWLEDGEMENTS

This research was supported by the Bureau of Standards, Metrology and Inspection, Ministry of Economic Affairs, and the Industrial Technology Research Institute. Additional support was provided by the National Science and Technology Council under Grants.

DECLARATIONS

Conflict of interest: The authors declare no conflict of interest.

Data availability: The source-level repeat data and formula-linked audit workbook are not publicly available because they contain confidential industrial measurement data, supplier information, and internal-company information. The processed and anonymized results needed to support the findings are included in this article.

REFERENCES

Arenhart, R.S., Pizzolato, M., Menin, P.L., and Hoch, L., "Devices for Interim Check of Coordinate Measuring Machines: A Systematic Review," *MAPAN*, Vol. 36, No. 1, pp. 157–173 (2021).

Bell, S.A., *A Beginner's Guide to Uncertainty of Measurement* (2001).

Buchta, C., Benka, B., Delatour, V., Faé, I., Griesmacher, A., Hellbert, K., Huggett, J., Kaiser, P., Kammel, M., Kessler, A., Kessler, H.H., Müller, D., Rosendahl, J., Scheiblauer, H., Schweiger, C.R., Zeichhardt, H., and Cobbaert, C.M., "Reference, Calibration and Referral Laboratories—A Look at Current European Provisions and Beyond," *Clinical Chemistry and Laboratory Medicine*, Vol. 63, No. 4, pp. 656–669 (2025).

Burdick, R.K., Borror, C.M., and Montgomery, D.C., "A Review of Methods for Measurement Systems Capability Analysis," *Journal of Quality Technology*, Vol. 35, No. 4, pp. 342–354 (2003).

Chen, J.R., Hung, S.H., and Lin, M.X., "Comparison of Metrology-Grade Squareness Error Measurement Methods," *Journal of the Chinese Society of Mechanical Engineers*, Vol. 46, No. 1, pp. 33–40 (2025).

Fesperman, R.R., Moylan, S.P., and Donmez, M.A., "Methods, Practices, and Standards for Evaluating On-Machine Touch Trigger Probing of Workpieces," *Proc. of the 25th Annual Meeting of the American Society for Precision Engineering (ASPE)*, pp. 309–312 (2010).

Gawor, A., Kurek, E., Ruszczyńska, A., and Bulska, E., "Key Issues Related to the Accreditation of Academic Laboratories," *Accreditation and Quality Assurance*, Vol. 26, No. 6, pp. 285–291 (2021).

Hansen, H.N., and De Chiffre, L., "An Industrial Comparison of Coordinate Measuring Machines in Scandinavia with Focus on Uncertainty Statements," *Precision Engineering*, Vol. 23, No. 3, pp. 185–195 (1999).

Hsieh, T.H., Lin, M.X., and Yeh, K.T., "Measuring Parametric and Volumetric Errors in a Four-Axis CMM Using a Hole Plate," *International Journal of Precision Engineering and Manufacturing*, Vol. 25, No. 5, pp. 959–979 (2024).

Hsieh, T.H., Lin, M.X., and Watanabe, T., "Separation of All Motion Errors in a Rotary Motor Using Dual Rotary Encoders," *International Journal of Precision Engineering and Manufacturing*, Vol. 26, No. 1, pp. 117–131 (2025).

Hsu, C.H., Chen, J.R., Hsu, F.H., and Chen, Y.T., "A Novel Measurement Method for Determining Geometric Errors of Rotary Tables by Using LaserTRACER and Reflectors," *Applied Sciences*, Vol. 13, No. 4, Article No. 2419 (2023).

ISO/IEC Guide 98-3:2008, Uncertainty of Measurement—Part 3: Guide to the Expression of Uncertainty in Measurement (2008).

Jakubiec, W., Płowucha, W., and Starczak, M., "Analytical Estimation of Coordinate Measurement Uncertainty," *Measurement*, Vol. 45, No. 10, pp. 2299–2308 (2012).

Lin, M.X., and Hsieh, T.H., "Geometric Error Parameterization of a CMM via Calibrated Hole Plate Archived Utilizing DCC Formatting," *Applied Sciences*, Vol. 13, No. 10, Article No. 6344 (2023).

Petrò, S., and Moroni, G., "A Statistical Point of View on the ISO 10360 Series of Standards for Coordinate Measuring Systems Verification," *Measurement*, Vol. 172, Article No. 108937 (2021).

Rahman, M.M., and Mayer, R., "Performance of a Five-Axis Machine Tool as a Coordinate Measuring Machine (CMM)," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, Vol. 10, No. 5 (2016).

Šafarič, J., Klobučar, R., and Ačko, B., "Measurement Setup and Procedure for Precise Step Gauge Calibration," *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, pp. 1–10 (2021).

Sato, O., Takatsuji, T., Matsuzaki, K., Watanabe, M., Kajima, M., Miura, Y., and Nakanishi, S., "Practical Experimental Design and Uncertainty Evaluation Method for Dimensional and Form Measurements Using Coordinate Measuring Machines," *Measurement*, Vol. 227, Article No. 114224 (2024).

Schwenke, H., Knapp, W., Haitjema, H., Weckenmann, A., Schmitt, R., and Delbressine, F., "Geometric Error Measurement and Compensation of Machines—An Update," *CIRP Annals*, Vol. 57, No. 2, pp. 660–675 (2008).

Štrbac, B., Radlovački, V., Spasić-Jokić, V., Delić, M., and Hadžistević, M., "The Difference Between GUM and ISO/TC 15530-3 Method to Evaluate the Measurement Uncertainty of Flatness by a CMM," *MAPAN*, Vol. 32, No. 4, pp. 251–257 (2017).

Takatsuji, T., Osawa, S., and Kurosawa, T., "Uncertainty Analysis of Calibration of Geometrical Gauges," *Precision Engineering*, Vol. 26, No. 1, pp. 24–29 (2002).

Vardeman, S.B., and VanValkenburg, E.S., "Two-Way Random-Effects Analyses and Gauge R&R Studies," *Technometrics*, Vol. 41, No. 3, pp. 202–211 (1999).