https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


# Evaluating data quality and reference instrument robustness: insights from 12 years DI magnetometer comparisons in the Geomagnetic Network of China

Yufei He[1], Qi Li[1], Xudong Zhao[1], Suqin Zhang[1] and Fuxi Yang[2]

[1]Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China
[2]Earthquake Bureau of Xinjiang Province, Urumqi, 830011, China

*Correspondence to*: Qi Li (darcyli@163.com)

**Abstract.** A statistical analysis was conducted on 12 years of geomagnetic instrument comparison data from the Chinese Geomagnetic Network (GNC). The study reveals that when applying a 90% threshold, the corresponding instrument deviation thresholds are 0.21′ (D component) and 0.11′ (I component), which can serve as evaluation criteria at the network level. By integrating multi source uncertainty decomposition (including instrument errors, operator dependent errors, and pillar correction errors) and weighted ensemble analysis, systematic deviations between reference fluxgate theodolites and test instruments were quantified. Results demonstrate that reference instruments exhibit high stability and reliability, with mean deviations of -0.004′ (D) and 0.022′ (I), both within the 95% confidence interval, and no long term drift was observed. Operator dependent errors were successfully isolated, with 0.13′ (D) and 0.06′ (I), consistent with observed experimental findings, confirming that operator dependent errors constitute the primary contributor to instrument deviations. Notably, operator dependent errors in D are significantly higher than I due to the complexity of azimuth alignment. These findings highlight the critical role of instrument comparisons in effectively monitoring equipment performance and assessing observational quality across stations, while validating the feasibility of standardized instrument evaluation methods and the long term stability of reference instruments. Future efforts should integrate sensors and automation technologies to minimize human errors, thereby providing a higher quality data foundation for geophysical studies.

## 1 Intruduction

In geomagnetic observatories, variometers are employed to record continuous variations of the geomagnetic field. These variations are subsequently converted into absolute geomagnetic field values through the addition of baseline values derived from absolute measurements (Jankowski and Sucksdorff, 1996). This calibration process renders absolute measurements critical for ensuring the quality of continuous absolute geomagnetic data. However, inter observatory discrepancies in absolute instruments necessitate systematic instrument comparisons as an essential component of modern geomagnetic observation systems.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


Contemporary absolute measurements primarily utilize two high precision instruments: (i) fluxgate theodolites (designated as Declination Inclination Magnetometers, DIMs) for measuring declination (D) and inclination (I), and (ii) proton magnetometers for total field intensity (F) determinations. While technological advancements have reduced the required frequency of instrument comparisons, such calibrations remain indispensable for maintaining high quality geomagnetic datasets (Zhang et al., 2024). To standardize global geomagnetic observations, the IAGA Division V Working Group V-OBS has successfully organized over twenty biennial international instrument comparison sessions to date (Loubser et al., 2002; Masami et al., 2004, Reda et al., 2007; Love et al.,2009; He et al., 2011; Hejda et al.,2013; and so on). In China, the Geomagnetic Network Center (GNC) integrates these comparisons into its quality control framework, serving dual roles as both data hub and quality assurance authority for national observatories (Li et al., 2012; Zhang et al., 2016).

Since the digital transformation of Chinese geomagnetic observatories, the GNC has implemented successive generations of DIMs, including: Chinese CTM-DIM,Hungarian MINGEO-DIM,British MAG-01-DIM,Chinese GEO-DIM, Chinese TDJ2E-NM-DIM and so on. Notably, the manual operation inherent to these DIM systems introduces two critical uncertainty sources: i) inter instrument systematic biases, and ii) operator dependent operational variances. To mitigate these effects and unify observational standards across the network, the GNC has conducted multi annual comparative measurements and accumulated corresponding datasets since the digital transformation of GNC (He et al., 2019b). An instrument demonstrating high measurement accuracy and operational stability is typically designated as the national reference standard for GNC. Through systematic comparative measurements, each observatory's instruments are calibrated against this reference standard to quantify instrumental deviations, thereby achieving nationwide standardization of geomagnetic absolute observations. As the reference instrument, it requires not only meticulous routine maintenance but also periodic metrological verification to ensure sustained measurement precision.

This study investigates the metrological robustness of the reference instrument through historical comparative measurement data. The methodology initiates with an overview of the comparative measurement protocol and multi year calibration results. Subsequently, error propagation analysis is conducted based on the measurement framework and longitudinal datasets, yielding uncertainty estimates for each comparative session. Concurrently, systematic deviations of the reference instrument relative to iterative measurements are quantified to evaluate its long term stability.

## 2 Measurement and comparative methodology

### 2.1 Measurement principles

The geomagnetic field, being a vector quantity, requires precise determination of both magnitude and directional components. The DIM, comprising a theodolite integrated with a fluxgate sensor, serves as the standard instrument for determining geomagnetic field direction (declination D and inclination I). The fluxgate sensor, mounted coaxially with the theodolite's optical axis, operates on the null detection principle: it generates zero output when aligned perpendicular to the geomagnetic field vector. Directional determination is achieved by identifying sensor null positions, with angular coordinates recorded via

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


theodolite circle readings. The angular between two directions can be determined by computing the difference between their respective readings on the instrument's horizontal circle. This is the fundamental principle of theodolite angle measurement.

### 2.2 Declination and inclination measurement

Geomagnetic declination and inclination measurements are performed within the horizontal and vertical planes of the theodolite respectively. The declination determination involves two sequential operations: establishing true north orientation and identifying the geomagnetic meridian direction. The true north orientation is calibrated by aligning the telescope's optical axis with a predefined azimuth marker (Fig.1), with the horizontal circle reading recorded as M. As the azimuth value A of the marker given, the true north position can be calculated. Subsequently, the geomagnetic meridian direction is identified by searching the fluxgate null position in the horizontal plane (the vertical circle maintain at 90 °or 270 °) and recording horizontal reading (D'). The geomagnetic declination D is then derived from the differential angular measurement following the formula:

$$D = D' - M + A \quad (1)$$

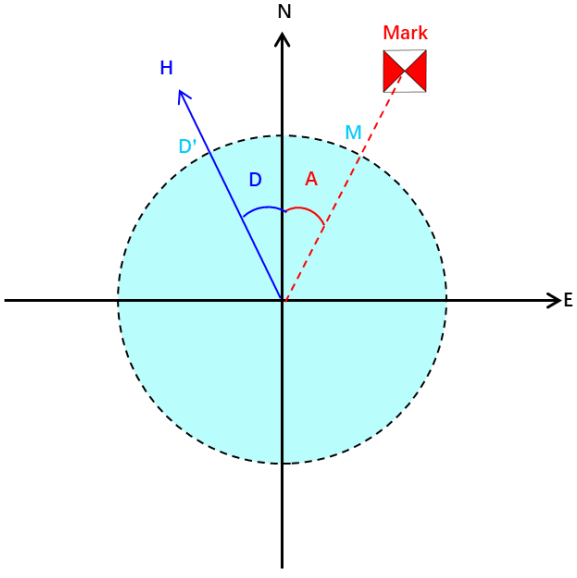


**Figure 1: Measurement principle of the declination.**

Inclination measurement follows analogous procedures in the vertical plane, omitting azimuth marker referencing. The fundamental equation remains consistent with declination calculation.

### 2.3 Error mitigation strategy

A four position calibration protocol is implemented in field observations to mitigate instrumental errors, requiring sequential measurements at distinct telescope orientations (Bitterly et al., 1984). The measurement procedure followed the guide published by IAGA (Newitt et al., 1996). The declination measurement protocol involves four configurations: (i) telescope

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


East/sensor up ($D_1$), (ii) telescope West/sensor down ($D_2$), (iii) telescope East/sensor down ($D_3$), and (iv) telescope West/sensor up ($D_4$). The final declination value is derived through arithmetic averaging:

$$\overline{D'} = (D_1 + D_2 + D_3 + D_4)/4 \quad (2)$$

An analogous procedure governs inclination measurement, with positional configurations: (i) telescope North/sensor up ($I_1$), (ii) telescope South/sensor down ($I_2$), (iii) telescope North/sensor down ($I_3$), and (iv) telescope South/sensor up ($I_4$). The inclination is calculated as:

$$I = (I_1 + I_2 - I_3 - I_4)/4 + 90 \quad (3)$$

This methodology effectively compensates for fluxgate sensor optical axis misalignment (Deng et al., 2010). Two distinct circle reading techniques are employed: the null method (exact zero point detection) and the offset method (near zero linear region utilization). As demonstrated by Xin (2003), Lu (2008), and Deng (2011), modern theodolites' high output linearity enables equivalent accuracy between methods, even with minor operator induced magnetic interference, making the offset method preferable for operational efficiency. Integration of declination (D), inclination (I), and proton magnetometer derived total intensity (F) enables precise determination of the absolute geomagnetic vector, facilitating subsequent variometer baseline calculations across all components.

### 2.4 Baseline comparison methodology

Ensuring high quality geomagnetic observations necessitates systematic verification of inter instrument discrepancies across observatories. The comparative analysis of absolute measurement instruments constitutes an essential quality assurance measure in geomagnetic monitoring networks. Practical constraints, including limited pillar availability, multiple DIMs, and operator skills proficiency, render synchronous multi instrument comparisons operationally challenging. Modern variometers exhibit high precision performance with quasi constant baseline characteristics under stable operation. Current comparison protocols therefore employ independent absolute measurements followed by baseline value cross validation. Zhang's (2010) empirical validation (08:30-16:30 LT) confirmed baseline stability remains unaffected by geomagnetic field fluctuations during calibration windows, establishing baseline comparison as a methodologically robust approach. The baseline value computation for component W follows the standardized formulation in the INTERMAGNET reference manual (St-Louis, 2024):

$$W_B(k) = W_o(i{:}j) - W_R(k) \quad (4)$$

Were, $W_o(i{:}j)$ is the absolute field mean over interval *i-j* (typically minutes), $W_R(k)$ is the variometer recorded value at time $k$, and $W_B(k)$ is the derived baseline value.

When absolute measurements are performed across distinct pillars, baseline correction to the reference pillar requires incorporation of inter pillar discrepancies. The generalized formulation for component W correction is:

$$\Delta U_W(s,o) = W_B(s) - W_B(o) + \Delta W_{so} \quad (5)$$

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


Where, $s$ is the reference pillar designation, o is the non-reference observation pillar, $W_B(s)$ and $W_B(o)$ are respectively the baseline values from reference and observation pillars, $\Delta W_{so}$ is the predetermined inter pillar correction term, $\Delta U_W(s,o)$ is the final instrument discrepancy metric.

This methodology enables standardized cross observatory fluxgate theodolite comparisons through pillar referenced baseline alignment. The obtained difference values $\Delta U_W(s,o)$ provide quantitative evaluation parameters for assessing absolute observation data quality across participating instruments.

### 2.5 GNC DIMs intercomparison results

Based on the aforementioned methodology, the GNC has completed the comparative measurements of DIMs across the entire observatory network from 2010 to 2024. Table 1 presents the timeline and locations of these comparative measurements, while Table 2 lists station codes, station names, and corresponding instrument types.

**Table 1: The list of DIM intercomparisons of GNC**

| No. | Date | Year | Observatory |
|---|---|---|---|
| 1 | 14–17 September | 2010 | Changchun |
| | 10–12 September | | Qianling |
| 2 | 17–19 October | 2012 | Dalian |
| | 24–26 October | | Shaoyang |
| 3 | 21–26 October | 2014 | Qianling |
| 4 | 15–20 September | 2015 | Urumqi |
| 5 | 21–29 September | 2016 | Qianling |
| 6 | 12–20 September | 2017 | Urumqi |
| 7 | 10–18 September | 2018 | Hongshan |
| 8 | 17–25 September | 2019 | Qianling |
| 9 | 14–21 October | 2020 | Qianling |
| 10 | 25–29 September | 2022 | Hongshan |
| 11 | 26 July – 1 August | 2023 | Urumqi |
| 12 | 19–25 September | 2024 | Qianling |

**Table 2: The list of instrument type and its observatory**

| No. | Observatory | Instrumnent | No. | Observatory | Instrumnent | No. | Observatory | Instrumnent |
|---|---|---|---|---|---|---|---|---|
| 0 | GNC | Mingeo | 15 | Kashi | Mingeo | 31 | Taiyuan | Mingeo |
| 1 | Beijing | Mingeo | 16 | Lhasa | Mingeo | 32 | Taian | GEO-DI |

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | Changli | Mag01 | 17 | Lanzhou | Mingeo | 32 | Taian | CTM |
| 2 | Changli | CTM | 18 | Lijiang | Mag01 | 33 | Tianshui | Mag01 |
| 3 | Chengdu | Mingeo | 19 | Luoyang | Mingeo | 34 | Tonghai | Mingeo |
| 3 | Chengdu | Mag01 | 20 | Malingshang | Mingeo | 35 | Wujiahe | Mingeo |
| 4 | Chongming | Mag01 | 21 | Manzhouli | Mingeo | 36 | Urumqi | Mingeo |
| 4 | Chongming | TDJ2E-NM | 21 | Manzhouli | CTM | 37 | Wuhan | Mingeo |
| 5 | Dalian | Mingeo | 22 | Mengcheng | Mag01 | 37 | Wuhan | Mag01 |
| 6 | Dedu | Mingeo | 22 | Mengcheng | CTM | 37 | Wuhan | CTM |
| 7 | Enshi | Mingeo | 23 | Qianling | Mingeo | 38 | Xichang | CTM |
| 8 | Golmud | Mingeo | 23 | Qianling | Mag01 | 39 | Xilinhot | Mag01 |
| 9 | Guiyang | Mingeo | 23 | Qianling | CTM | 40 | Xiannushan | Mag01 |
| 10 | Hangzhou | CTM | 24 | Qiemo | Mag01 | 41 | Yinchuan | Mingeo |
| 10 | Hangzhou | Mag01 | 25 | Qiongzhong | Mingeo | 42 | YingCheng | Mingeo |
| 11 | Hongshan | Mag01 | 26 | Quanzhou | Mingeo | 43 | Yongning | Mingeo |
| 11 | Hongshan | TDJ2E-NM | 26 | Quanzhou | CTM | 44 | Yulin | CTM |
| 12 | Huhehot | CTM | 27 | Shaoyang | Mingeo | 45 | Changchun | Mingeo |
| 13 | Jiayuguan | Mingeo | 28 | Sheshan | Mingeo | 46 | Zhaoqing | Mingeo |
| 14 | Jinghai | Mingeo | 29 | Shexian | Mag01 | 46 | Zhaoqing | Mag01 |
| 14 | Jinghai | Mag01 | 30 | Shiquanhe | Mingeo | 46 | Zhaoqing | TDJ2E-NM |

Figures 2(a) and (b) respectively illustrate the instrumental deviations in declination ($\Delta U_D$) and inclination ($\Delta U_I$) between station instruments and reference standards. Colored dots in the figures represent measurement results from different years, with vertical coordinates indicating instrumental deviations, while the dot sizes indicate the standard deviations of
130 measurements, scaled according to the legend on the right. This graphical representation enables a comprehensive evaluation of observational data quality at both individual station and the network levels.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


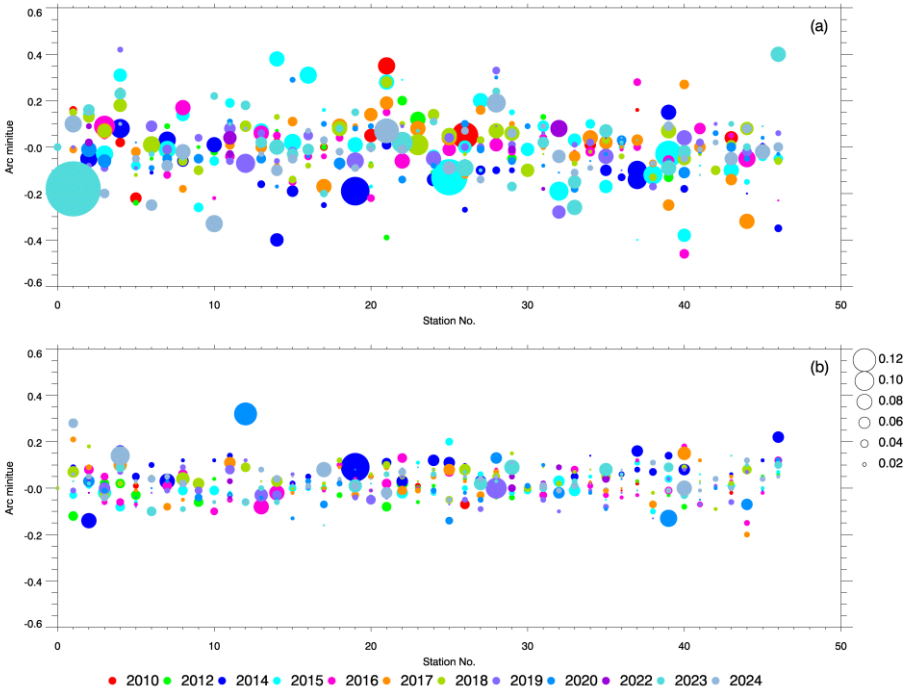


**Figure 2: Instrumental deviations in declination D (a) and inclination I (b). The dots represent the instrument deviations; the size of the dots is the standard deviation with the scale on the right. The colors indicate different years.**

135

First, the figures provide insights into instrument performance and operator proficiency. Small dots with large central values indicate significant instrumental deviations, suggesting potential instrument malfunctions or centering errors (e.g., misalignment) that may reflect technical shortcomings in observational practices. Conversely, large dots with small central values signify dispersed data, which could arise from instrument related issues (e.g., unclear optical paths affecting reading 140 accuracy) or inconsistent operational practices. Furthermore, the dispersion of multiple dots corresponding to the same station reflects the overall data quality at individual stations. Frequent personnel changes at some stations have introduced operator dependent errors, manifesting as increased dispersion. This graphical approach thus effectively monitors instrument performance and evaluates observational quality across stations.

Second, the collective distribution of all dots permits assessment of network wide data quality over multiple years. To quantify 145 network performance, we conducted statistical analyses of all instrumental deviations. Statistical analysis of all deviations (Fig. 3) reveals that declination (D) and inclination (I) measurements approximate normal distributions, with means of 0.00′ and 0.02′, and standard deviations of 0.13′ and 0.07′, respectively. Approximately 75.1% of declination and 86.8% of inclination measurements fall within ±1σ of the mean. When adopting the 90% thresholds as evaluation criteria for the entire network, the corresponding instrument deviations thresholds are 0.21′ for D and 0.11′ for I, indicating excellent consistency among 150 network fluxgate instruments. Notably, declination measurements exhibit greater dispersion than inclination values. This

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


discrepancy stems from the additional azimuth marker alignment required in declination measurements—a process more susceptible to operators compared to inclination measurements.

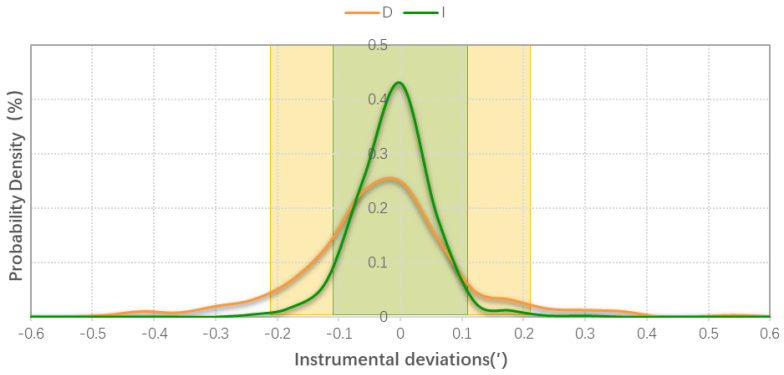


**Figure 3: The instrumental deviations of declination D (orange line) and inclination I (green line).**

This simplified comparative measurements provide an efficient mechanism for identifying inter station discrepancies, monitoring instrument performance, and ensuring standardized high quality observations across the network. However, the efficacy of this mechanism critically depends on the precision and accuracy of reference instruments. Beyond routine maintenance and calibration of reference standards, it is necessary to analyze its long term stability and reliability. The following section will evaluate the reference instruments using all comparative measurement data through uncertainty analysis.

## 3 Robustness evaluation of the reference instrument

The robustness evaluation of the reference instrument requires quantification of its systematic deviation from true values and associated uncertainties. The assessment involves three key components: determination of true values, characterization of the reference instrument's measurements, and statistical analysis of deviations. This study leverages over 12 years intercomparison data to conduct the evaluation. First, the true values and uncertainties are assessed for each intercomparison session, followed by analyzing deviations between the reference instrument and true values. Second, a temporal deviation model is established using multi years intercomparison results to evaluate the long term stability of the reference instrument. Finally, annual composite uncertainties are calculated by integrating multi years observational means and standard deviations from participating instruments, reflecting the dynamic characteristics of the measurement system. A comprehensive evaluation of the reference instrument's robustness is achieved through time series analysis and multi sources uncertainty synthesis.

### 3.1 Uncertainty analysis method for a single intercomparison

During the intercomparison process, fluxgate theodolites are employed to observe magnetic declination and inclination, with baseline value comparisons serving as the final validation step. The analysis of uncertainties in this process encompasses

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


several key error sources, including instrumental intrinsic errors, repeatability errors, operator induced errors, and pillar correction errors. Environmental interference is excluded from consideration due to controlled laboratory conditions.

The fluxgate theodolite consists of two primary components: the theodolite and the fluxgate sensor. For the theodolite component, instrumental intrinsic errors can be evaluated based on national standards and technical specifications. As shown in Table 3, the maximum permissible single set standard deviations for horizontal and vertical angles vary across instrument classes (AQSIQ et al.,2015). Assuming a normal distribution for these deviations, the Type B standard uncertainty for the theodolite is calculated using Eq. (6):

$$u_{b1} = \Delta/3 \tag{6}$$

**Table 3: Maximum Permissible Single Set Standard Deviations for Theodolites**

| Class | Horizontal Angle | Vertical Angle |
|---|---|---|
| **DJ07** | 0.6″ | 2.0″ |
| **DJ1** | 0.8″ | 2.0″ |
| **DJ2** | 1.6″ | 6.0″ |
| **DJ6** | 4.0″ | 10.0″ |

For the fluxgate sensor, errors arise from the liquid crystal display's limited resolution ($\varepsilon$=0.1 nT), which follows a uniform distribution. The corresponding Type B standard uncertainty is derived via Eq. (7):

$$u_{b2} = \varepsilon/\sqrt{3} \tag{7}$$

The synthesized intrinsic uncertainty for each instrument is then computed as the root sum square of these components (Eq. 8):

$$u_{inst,i} = \sqrt{u_{b1,i}^2 + u_{b2,i}^2} \tag{8}$$

Repeatability and operator related uncertainties are addressed through systematic data collection. Each operator instrument combination performs 6~8 repeated measurements, enabling the calculation of repeatability standard deviations ($s_{rep,i}$) and associated Type A uncertainties ($u_{rep,i}$) using Eq. (9):

$$u_{rep,i} = \frac{s_{rep,i}}{\sqrt{N}} \tag{9}$$

$$s_{rep,i} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(x_{i,k} - \frac{1}{N}\sum_{k=1}^{N} x_{i,k}\right)^2} \tag{10}$$

Operator induced errors are decoupled from instrumental effects by analyzing inter station variances. The between group standard deviation ($s_{between}$) is calculated across multiple operator instrument combinations following Eq. (11), and operator uncertainty ($u_{oper}$) is derived by subtracting the averaged instrumental uncertainties thought Eq. (12):

$$s_{between} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\overline{x_i} - \bar{\bar{x}})^2}\ , \qquad \bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\overline{x_i} \tag{11}$$

$$u_{oper} = \sqrt{s_{between}^2 - \frac{1}{N}\sum_{i=1}^{N} u_{inst,i}^2} \tag{12}$$

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


Pillar correction errors ($u_{pier}$) are incorporated based on station specific deviation measurements. The total synthesized uncertainty for each instrument is then aggregated as the root sum square of all contributing factors, according to Eq. (13):

$$u_i = \sqrt{u_{inst,i}^2 + u_{oper}^2 + u_{rep,i}^2 + u_{pier,i}^2} \quad (13)$$

Finally, the ensemble mean ($\mu_{group}$) and combined uncertainty ($u_{group}$) for all instruments are computed using a weighted average approach. Weights ($\omega_i$) are assigned inversely proportional to the square of each instrument's total uncertainty, ensuring higher precision instruments exert greater influence, using Eq. (14) and (15):

$$\mu_{group} = \frac{\sum_{i=1}^{N} \omega_i \mu_i}{\sum_{i}^{N} \omega_i} \quad , \quad \omega_i = \frac{1}{u_i^2} \quad (14)$$

$$u_{group} = \frac{1}{\sqrt{\sum_{i}^{N} \omega_i}} \quad (15)$$

This comprehensive methodology transforms the intercomparison into a robust experiment integrating multi operators' collaboration, parallel instrumentation, and repeated measurements, ensuring rigorous uncertainty quantification.

### 3.2 Multi years intercomparison analysis method

Building on the uncertainty analysis of single intercomparison sessions, this section evaluates the long term stability and robustness of the reference instrument using data accumulated over 12 years of digitized observations within the GNC. Each intercomparison session involves comparing the reference instrument, mounted on a standardized pillar, against the ensemble results of participating instruments. Since the reference instrument requires no pillar correction, its mean value ($\mu_s$) and associated uncertainty ($u_{rep,s}$) are derived from 6~8 repeated measurements per session. The repeatability standard deviation ($s_{rep,s}$) and corresponding uncertainty ($u_{rep,s}$) are calculated using Eq. (16) and (17):

$$u_{rep,s} = \frac{s_{rep,s}}{\sqrt{N}} \quad (16)$$

$$s_{rep,s} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(x_{s,k} - \mu_s\right)^2} \quad , \quad \mu_s = \frac{1}{N}\sum_{k=1}^{N} x_{s,k} \quad (17)$$

The total uncertainty of the reference instrument ($u_s$) incorporates its intrinsic error ($u_{inst,s}$) operator induced uncertainty ($u_{oper}$), and repeatability uncertainty, as expressed in Eq. (18):

$$u_s = \sqrt{u_{inst,s}^2 + u_{oper}^2 + u_{rep,s}^2} \quad (18)$$

The deviation ($\Delta$) between the reference instrument and the weighted ensemble mean ($\mu_{group}$) of all participating instruments, along with its uncertainty ($u_\Delta$), is quantified for each session using Eq. (19) and (20):

$$\Delta = \mu_{group} - \mu_s \quad (19)$$

$$u_\Delta = \sqrt{u_s^2 + u_{group}^2} \quad (20)$$

A consistency criterion ($|\Delta| \leq 2u_\Delta$) is applied to verify agreement at a 95% confidence level.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


To assess long term stability, deviations (Δ) from $M$ years intercomparison sessions are compiled into a time series plot, enabling visual detection of potential drifts caused by environmental fluctuations or instrumental aging. The multi years mean deviation ($\overline{\Delta}$) and its uncertainty ($u_{\overline{\Delta}}$) are calculated through Eq. (21) and (22):

$$\overline{\Delta} = \frac{1}{M}\sum_{m=1}^{M}\Delta_m \tag{21}$$

$$u_{\overline{\Delta}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}u_{\Delta,m}^2 + \left(\frac{s_\Delta}{\sqrt{M}}\right)^2} \tag{22}$$

Here, ($s_\Delta$), represents the standard deviation of Δ across all sessions. The final robustness criterion ($|\overline{\Delta}| \le 2u_{\overline{\Delta}}$) ensures the reference instrument's performance remains within acceptable bounds over extended periods. This integrated approach combines temporal trend analysis with uncertainty propagation, providing a comprehensive evaluation framework for maintaining measurement integrity in long term geomagnetic monitoring.

### 3.3 Results analysis

Using the methodology described above, we analyzed 12 years of instrumental deviation data for declination (D) and inclination (I). The time series of mean differences (Δ) between the reference instrument and all tested instruments, along with their uncertainties ($u_\Delta$), are shown in Fig. 4. In the figure, orange and green histograms represent mean differences for D and I, respectively, while curves of corresponding colors indicate twice the uncertainty ($2u_\Delta$). According to the criterion $|\overline{\Delta}| \le 2u_{\overline{\Delta}}$, most mean differences fall within the $2u_\Delta$ range.

However, mean differences both D and I exceeded this threshold in 2014. A retrospective review of the raw data revealed no definitive cause for this anomaly. Notably, 2014 involved an observational training program where measurements were conducted by inexperienced personnel, and the definite uncertainty was notably low. This suggests potential transient impacts on the reference instrument. Additionally, the mean difference for D in 2018 slightly exceeded $2u_\Delta$, though no conclusive explanation has been identified. Nevertheless, the long term mean difference data demonstrate that the reference instrument has maintained stable operation and reliable performance throughout the study period.

Finally, using Eq. (21) and (22), we evaluated the reference instrument's long term stability by calculating the multi years average mean differences and their uncertainties. For declination (D), the average mean difference was $\overline{\Delta_D} = -0.004'$ with an uncertainty of $u_{\overline{\Delta D}} = 0.054'$. For inclination (I), the values were $\overline{\Delta_I} = 0.022'$ and $u_{\overline{\Delta I}} = 0.023'$. Applying the criterion $|\overline{\Delta}| \le 2u_{\overline{\Delta}}$ (95% confidence level), both D and I meet this requirement, confirming the reliability of the reference instrument's long term observational data.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


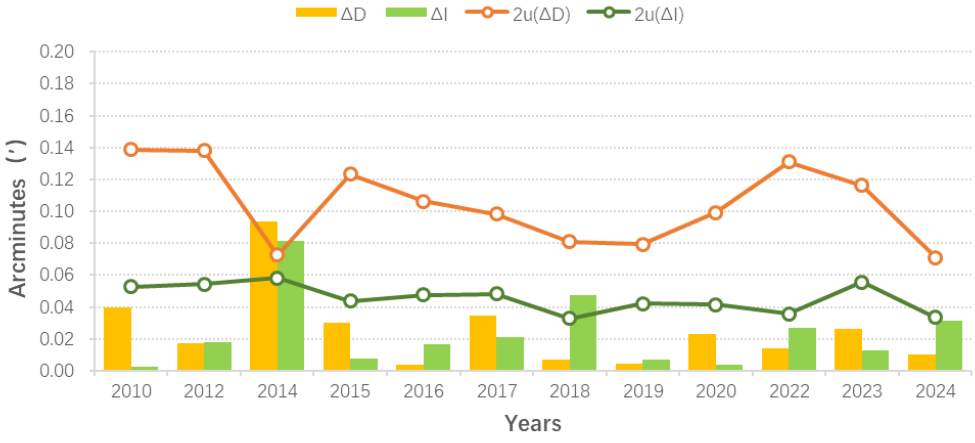


**Figure 4: Time Series of Mean Differences and Uncertainties Between Reference and Tested Instruments.**

Operator dependent errors (human discrepancies) during comparative measurements were calculated using Eq. (12), as illustrated in Fig. 5. Orange and green dots represent discrepancies for D and I, respectively. Results show consistently higher operator dependent errors in declination measurements compared to inclination, with D errors persistently exceeding I values. This discrepancy arises from the additional azimuth marker alignment step required for declination measurements, which introduces greater operator variability. The mean operator dependent errors were 0.13′ for D and 0.06′ for I, aligning closely with experimental results（0.18′ for D and 0.08′ for I）reported by He (2019a), thereby validating the methodology's effectiveness in quantifying human induced errors.

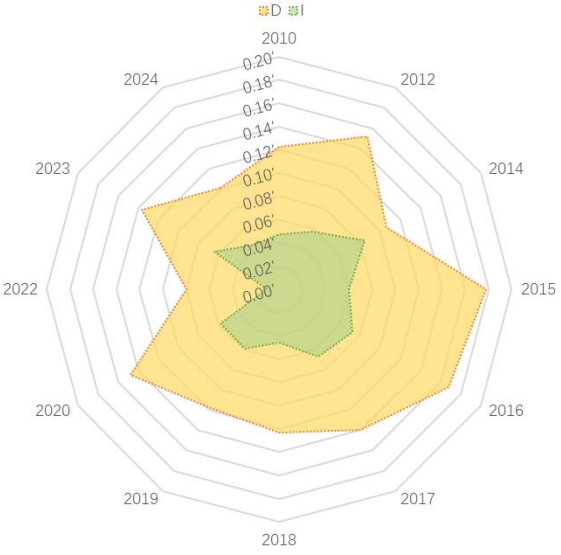


**Figure 5: Operator dependent Errors in Declination (D) and Inclination (I) Measurements.**

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


## 4 Conclusion and outlook

This study systematically analyzed 12 years of comparative measurement data from the GNC. The results underscore the critical importance of instrument comparisons in geomagnetic observatory networks, demonstrating their dual role in (i) monitoring instrument performance and evaluating data quality at both station and network scales, and (ii) providing researchers with transparent insights into the accuracy of absolute measurements. These assessments also offer valuable references for evaluating the reliability of scientific conclusions derived from geomagnetic data.

A comprehensive evaluation framework was developed to assess the robustness of reference instruments. By decomposing uncertainties into instrument specific errors, repeatability errors, operator dependent discrepancies, and pillar correction errors during individual comparisons, we achieved precise quantification of multi sources uncertainties. The weighted mean method enhances result reliability by prioritizing high precision instruments. The integration of multi instruments datasets through weighted averaging significantly improved result reliability. By constructing time series of mean differences (Δ) and their uncertainties between the reference instrument and tested instruments, we analyzed the long term stability of the reference instrument.

Furthermore, multi years data analysis revealed the mean differences and interannual composite uncertainties, while significance tests based on mean differences and expanded uncertainties provided objective criteria for instrument performance, ensuring measurement consistency in complex environments. Additionally, our methodology successfully separated operator dependent errors from inherent instrument errors, with computational results aligning closely with experimental validations. However, this approach has limitations. For instance, it overlooks operator instrument interactions and environmental factors (e.g., humidity fluctuations), which may introduce systemic biases. Long term stability analysis also requires extensive multi years comparative data to ensure statistical power, limiting rapid field applications. Future research should focus on model optimization, such as incorporating environmental sensor data to establish temperature/humidity compensation mechanisms or developing automated tools to streamline multi sources uncertainty synthesis. With continuous refinement, this methodology holds promise for advancing standardization and long term stability in geomagnetic observation networks.

Furthermore, geomagnetic observatories serve as primary facilities for measuring the secular variation of the Earth's magnetic field. The measurement accuracy for directional elements (e.g., declination and inclination) is typically required to be below 0.1', while the accuracy for intensity elements (e.g., total field strength) should be within 1nT. Modern instruments, such as the Zeiss 010B fluxgate theodolite, theoretically possess sufficient precision to achieve these targets. However, in practice, attaining such accuracy remains challenging due to various sources of error, particularly operator dependent errors. With advancements in automation and the global adoption of high precision instruments (Rasson et al.,2011; Gonsette et al., 2017; Hegymegi et al.,2017; and so on) in the future, it is anticipated that operator dependent discrepancies will be eliminated, thereby obtaining higher quality geomagnetic observational data.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


**Data availability**

The raw data are available upon request from the corresponding author at darcyli@163.com.

**Author contribution**

YH and QL initiated the study and designed the analysis methods. XZ and FY carried them out. SZ analyzed the data and results. YH prepared the manuscript with contributions from all coauthors.

**Competing Interests**

The authors have no competing interests to declare.

**Acknowledgements**

The authors extend sincere gratitude to all colleagues involved in the instrument comparisons, whose dedicated efforts were pivotal to the realization of this research.

**Funding Information**

Supported by National Key R&D Program of China (2023YFC3007404); National Natural Science Foundation of China (42374092); DI Magnetometer Comparison (0525205).

**References**

Bitterly,J.,J., Cantin, M., Schlich, R., Folques,J., and Gilbert D.: Portable magnetometer theodolite with fluxgate sensor for earth's magnetic field component measurements, Geopysical Surveys, 6, 233–239, 1984.

Deng N., Yang, D. M., Yang, Y. F., and Chen, J.: On problems of absolute measurement in geomagnetic observatory, Journal of Geodesy and Geodynamics, 30(A01), 129–134, 2010.

Deng, N., Yang, D. M., He, Y. F., and Yang, Y. F.: Study on the applicability of offset method at readings of ten minutes of arc in geomagnetic absolute measurement, Seismological and Geomagnetic Observation and Research, 32, 31–33, DOI:10.3969/j.issn.1003-3246.2011.02.006, 2011.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China (AQSIQ), Standardization Administration of China (SAC): Optical Theodolite, GB/T 3161-2015 , https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=C7878EF1AECE7FCF9D88EDE962F250E3 (last access: 21 May 2025), 2015.

Gonsette A, Rasson J, Bracke S, Poncelet, A., Hendrickx, O., and Humbled, F.: Fog-based automatic true north detection for absolute magnetic declination measurement, Geosci. Instrum. Method. Data Syst., 6(2), 439–446, 2017.

Hegymegi, L., Szöllősy, J., Hegymegi, C. and Domján, A.: Measurement experiences with FluxSet digital D/I station. Geosci. Instrum. Method. Data Syst., 6(2), 279–284, 2017.

He, Y. F., Yang, D. M., Zou, B. L., and Wang, J. J.: Report on the measurement session during the XIVth IAGA Workshop at Changchun magnetic observatory, Data Science Journal, 10, IAGA-02, https://www.jstage.jst.go.jp/article/dsj/10/0/10_IAGA-02/_article/-char/en, 2011.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


He, Y. F., Zhao, X. D., Wang, J. J., Yang, F. X., Li, X. J., Xin, C. J., Yan, W. S., and Tian, W. T.: The operator difference in absolute geomagnetic measurements, Geosci. Instrum. Method. Data Syst., 8, 21–27, 2019a.

He, Y. F., Zhao, X. D., Yang, D. M., Yang, F. X., Deng, N., and Li, X. J.: Analysis of Several Years of DI Magnetometer Comparison Results by the Geomagnetic Network of China and IAGA, Data Science Journal, 18, 1–11, 2019b.

Hejda, P., Chulliat, A., and Catalan, M.: Proceedings of the XVth IAGA Workshop on geomagnetic observatory instruments, data acquisition and processing, extended abstract volume, BOLETÍN ROA, No. 03/13, http://iaga_workshop_2012.roa.es/IAGA%20Extended%20Abstract%20Volume.pdf, 2013.

Jankowski, J., and Sucksdorff, C.: Absolute magnetic measurement, in: Guide for magnetic measurements and observatory practice, IAGA, Warszawa, Poland, 87–102, 1996.

Li, X. J., Yang, D. M., Zhang, S. Q., and He, Y. F.: The necessary of the manual observation in absolute observation, Seismological and Geomagnetic Observation and Research, 33(3), 201–205, 2012.

Loubser, L.: Proceedings of the Xth IAGA Workshop on geomagnetic instruments data acquisition and processing, Hermanus Magnetic Observatory, http://www.bgs.ac.uk/iaga/vobs/docs/XthIAGA_ws.pdf, 2002.

Love, J. J.: Proceedings of the XIIIth IAGA Workshop on geomagnetic observatory instruments, data acquisition and processing, USGS Open-File Report, 2009–1226, https://pubs.usgs.gov/of/2009/1226, 2009.

Lu, J. H., Wang, J. G., Li, X. Z., and Qiu, J.: Analysis and research on effects of absolute measurement time period on observation quality of geomagnetic baseline values, South China Journal of Seismology, 28, 113–119, DOI:10.13512/j.hndz.2008.04.003, 2008.

Masami, O., Takeshi, T., Katsuharu, K., et al.: Reports on the XIth IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition and Processing held at Kakioka/Tsukuba, Japan, in 2004, Technical Report of the Kakioka Magnetic Observatory, http://www.kakioka-jma.go.jp/publ/tr/2005/tr0008.pdf, 2005.

Newitt, L. R., Barton, C. E., and Bitterly, J.: Setting up equipment and taking measurements, in: Guide for magnetic repeat station surveys, IAGA, Warszawa, Poland, 43–45, 1996.

Rasson, J. L. and Gonsette, A.: The Mark II Automatic Diflux, Data Sci. J., 10, IAGA169–IAGA173, 2011.

Reda, J., and Neska, M.: Measurement session during the XII IAGA Workshop at Belsk, INST. GEOPHYS. POL. ACAD. SC., http://agp2.igf.edu.pl/agp/files/C-99/Podsumowanie.pdf, 2007.

St-Louis, B., INTERMAGNET Operations Committee and INTERMAGNET Executive Council: INTERMAGNET Technical Reference Manual, Version 5.1.1, (INTERMAGNET), Potsdam : GFZ Data Services, 134 p. https://doi.org/10.48440/INTERMAGNET.2024.001, Last accessed 20 May 2025.

Xin, C. J., Shen, W. R., Li, Q. H., and Tian, W. T.: The comparison and analysis of the baseline values of null method and offset method, Seismological and Geomagnetic Observation and Research, 24, 77–80, 2003.

Zhang, S. Q., and Yang, D. M.: Study on the stability and accuracy of baseline values measured during the calibrating time intervals, Data Science Journal, 10, IAGA19–IAGA24, 2011.

https://doi.org/10.5194/egusphere-2025-2406
Preprint. Discussion started: 16 June 2025


Zhang, S. Q., Fu, C. H., He, Y. F., Yang, D. M., Li, Q., Zhao, X. D., and Wang, J. J.: Quality control of observation data by the Geomagnetic Network of China, Data Science Journal, 15, 1–12, DOI:http://dx.doi.org/10.5334/dsj-2016-015, 2016.

Zhang, S. Q., Fu, C. H., Zhao, X. D., Zhang, X. X., He, Y. F., Li, Q., Chen, J., Wang, J. J., and Zhao, Q.: Strategies in the Quality Assurance of Geomagnetic Observation Data in China, Data Science Journal, 23, 1–11, DOI: https://doi.org/10.5334/dsj2024-009, 2024.