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

Yufei He[1,2], Xudong Zhao[1,2], Suqin Zhang[1], Qi Li[1] and Fuxi Yang[3]

[1]Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China
[2]Beijing Baijiatuan Earth Sciences National Observation and Research Station, Beijing 100095, China
[3]Earthquake Bureau of Xinjiang Province, Urumqi, 830011, China

Correspondence to: Xudong Zhao (zxd9801@163.com)

Abstract. An analysis was conducted on 12 sets of geomagnetic instrument comparison data from the Chinese Geomagnetic Network (GNC) between 2010 and 2024. First, by examining these comparison data, it was found that when their cumulative probabilities at the same level, the instrument differences for declination (D) are significantly higher than those for inclination (I). For the same set of instruments, as the frequency of observer changes increases, the instrument differences for D increase, while no significant change was observed for I. This indicates that inter-observer differences have a notable impact on D, primarily due to the complexity of aligning the azimuth marks and levelling instruments. Second, though a multi-source error uncertainty analysis, including instrument error, operator related error, pillar correction error and so on, the systematic differences between the reference fluxgate theodolite and the test instruments were quantified. The operator related errors of D and I were successfully separated and consistent with the observed experimental results, confirming that operator related error is the primary factor contributing to instrument differences. The analysis also validated the high stability and reliability of the reference instrument. The former finding can serve as an assessment criterion for network-level numerical quality, while the latter can be used to verify the long-term stability of the reference instrument.

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, the difference of absolute instruments between different observatories make systematic instrument comparisons as an essential component of modern geomagnetic observation systems.

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 Hungarian MINGEO-DIM, British MAG 01-DIM, Chinese CTM-DIM, Chinese GEO-DIM, Chinese TDJ2E-NM-DIM, etc. It should be noted that the manual operation inherent to these DIM systems introduces two critical uncertainty sources: i) inter instrument systematic biases, and ii) operator dependency operational variances. To mitigate these effects and unify observational standards across the geomagnetic network, the GNC has organized instrument (DIMs) comparison measurements almost annually since its digital transformation (He et al., 2019b). Through systematic comparison measurements, each observatory's instrument is calibrated against the reference instrument (DIMs), which is typically designated as the GNC standard and has the characteristics of high measurement accuracy and operational stability, to quantify instrumental differences and achieve nationwide standardization of absolute geomagnetic observations.

The following sections will first briefly introduce the measurement methods used in comparison measurements and the baseline based comparison methods. Then, the research will primarily focus on the historical comparison datasets. Firstly, statistical analysis will be conducted to explore valuable information from the statistical results. Secondly, uncertainty analysis method, widely used in experimental science, is applied to further quantify and extract relevant information from the datasets, to evaluated the uncertainty of the reference instrument relative to the observation results of all instruments in each comparison session. So the long term robustness of the reference instrument can be investigated based on multiyear results.

2 Measurement and comparison methodology

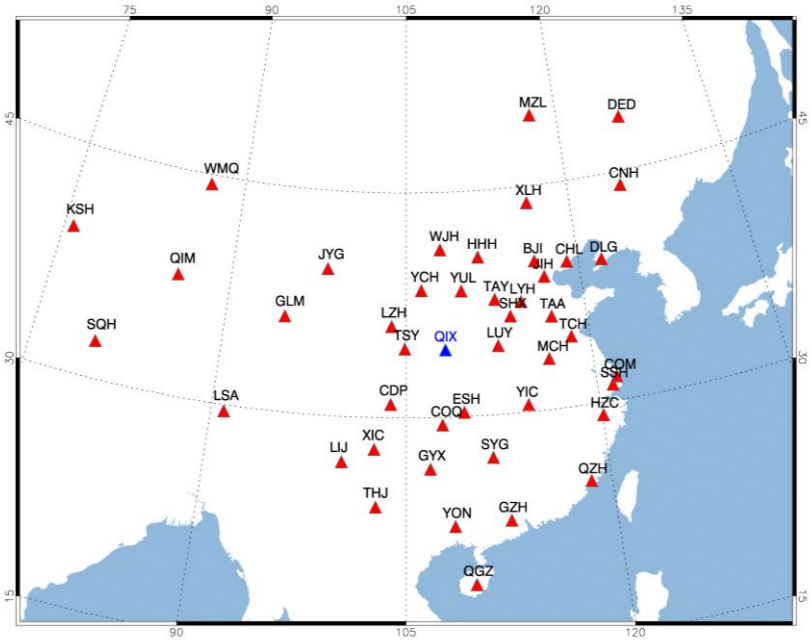
The geomagnetic field, being a vector quantity, requires precise determination of both magnitude and directional components. The DIM, comprising a theodolite integrated with a fluxgate sensor, serves as the standard instrument for determining geomagnetic field direction (declination D and inclination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

The fluxgate theodolites are high-precision instruments, but they inevitably contain certain errors, such as misalignment errors between the mechanical axis of theodolite, the optical axis of the telescope, and the magnetic axis of the fluxgate sensor; collimation errors; non-orthogonality errors of the horizontal and vertical axes; uneven graduation errors of the reading circle; index errors; and errors caused by non-zero electronic offsets, which prevent accurate determination of magnetic declination and inclination from a single reading of the horizontal/vertical circle (Lauridsen, 1985; Newitt et al., 1996; Csontos and Sugar, 2024). However, in theory, most of these errors can be eliminated through the four position measurement process, and some of them (two misalignment errors between the fluxgate sensor axis and the optical axis of the telescope in the horizontal/vertical planes, and the offset error of the fluxgate sensor) can be calculated from the measurement results (Bitterly et al., 1984). The specific measurement methods and procedure are described in Appendix A.1.

Ensuring high quality geomagnetic observations necessitates systematic verification of inter-instrument differences across observatories. The comparison 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 operating conditions, while underground observation rooms of geomagnetic observatories (far from cities or villages) can provide such operating conditions, including no influence of magnetic objects, low electromagnetic background noise, indoor annual temperature variation not exceeding 10 ℃, daily variation not exceeding 0.3 ℃, and so on. The current comparison protocols therefore employ independent absolute measurements followed by baseline value cross validation. The stability and accuracy of baseline values during a single calibration day were investigated by Zhang (2011), whose study confirmed that baseline values remained stable throughout the calibration period (8:30 to 16:30 local time), with geomagnetic activity exhibiting no significant impact on their accuracy. Consequently, direct comparison of baseline values is a valid approach for completing the analysis. The baseline calculation and instrument comparison methodology are presented in Appendix A.2.

3 Statistic analysis on comparison datasets

In each instrument comparison, all fluxgate theodolites instruments are brought to a specific observatory with excellent observation environment and compared with the reference instrument designated by the GNC. The reference instrument consists of a Zeiss 010B theodolite and a fluxgate sensor, with accuracies better than 1 arc second and 0.1 nT, respectively. The other instruments, known as testing instrument, come from 46 observatories of GNC, including five types as mentioned earlier. The locations of all observatories and the specific observatory (code QIX highlighted in blue) for comparison are shown in Fig.1. The codes of all the observatories and their corresponding instrument types, as well as the number of times each instrument participated in comparison and the number of operators (with non-repeated counts), are listed in Table 1. The No. 0 in Table 1 is the reference instrument, and the code represents GNC rather than the observatory. The relevant information and parameters for five types of fluxgate theodolites are provided in Table 2.



95 **Figure 1: The observatory locations of GNC.**

Table 1: The list of instrument type and its observatory

No.	Observatory Code	Instrument	Operators Number (non-repeated)	Comparison Frequency	No.	Observatory Code	Instrument	Operators Number (non-repeated)	Comparison Frequency
0	GNC	Mingeo	3	12	32	QIM	Mag-01H	2	9
1	BJI	Mingeo	3	8	33	QGZ	GEO	2	2
2	CHL	Mag-01H	3	10	34	QGZ	Mingeo	3	6
3	CHL	CTM	1	1	35	QZH	Mingeo	5	10
4	CDP	Mingeo	3	9	36	QZH	CTM	1	1
5	CDP	Mag-01H	1	1	37	SYG	Mingeo	3	10
6	CDP	GEO	2	2	38	SSH	Mingeo	2	9
7	COM	Mag-01H	4	5	39	SSH	TDJ2E-NM	1	2
8	COM	TDJ2E-NM	1	4	40	SHX	Mag-01H	3	10
9	DLG	Mingeo	4	10	41	SQH	Mingeo	2	3
10	DED	Mingeo	2	6	42	TAY	Mingeo	7	11
11	ESH	Mingeo	5	9	43	TAA	GEO-DI	2	4
12	GLM	Mingeo	3	8	44	TAA	CTM	1	1
13	GYX	Mingeo	3	5	45	TSY	Mag-01H	4	7
14	HZC	CTM	1	4	46	THJ	Mingeo	3	10
15	HZC	Mag-01H	1	1	47	WJH	Mingeo	4	10
16	LYH	Mag-01H	1	2	48	WMQ	Mingeo	3	10

17	LYH	TDJ2E-NM	3	5	49	WHN	Mingeo	5	6
18	HHH	CTM	2	2	50	WHN	Mag-01H	2	2
19	HHH	TDJ2E-NM	2	2	51	WHN	CTM	1	1
20	JYG	Mingeo	4	8	52	XIC	CTM	1	6
21	JIH	Mingeo	3	6	53	XLH	Mag-01H	3	6
22	JIH	Mag-01H	4	5	54	COQ	Mag-01H	5	7
23	KSH	Mingeo	4	9	55	YCH	Mingeo	2	8
24	LSA	Mingeo	5	9	56	YIC	Mingeo	2	4
25	LZH	Mingeo	5	11	57	YON	Mingeo	6	11
26	LIJ	Mag-01H	3	6	58	YUL	CTM	2	4
27	LUY	Mingeo	3	8	59	YUL	TDJ2E-NM	3	4
28	TCH	Mingeo	3	10	60	CNH	Mingeo	3	8
29	MZL	Mingeo	3	10	61	GZH	Mingeo	2	2
30	MCH	Mag-01H	3	9	62	GZH	Mag-01H	4	4
31	QIX	Mingeo	3	10	63	GZH	TDJ2E-NM	3	3

Table 2: The relevant information and parameters of the five fluxgate theodolites

Instrumnent	Theodolite			Sensor		
	Model	Resolution	Maximum permissible standard deviation	Model	Resolution	Offset
MinGeo-DIM	Theo 010A	1″, estimation 0.1"	$\leqslant \pm 2''$	Model G	0.1 nT	±1nT
Mag-01H	T1	6", estimation 3"	$\leqslant \pm 3''$	Mag A	0.1 nT	±1nT
TDJ2E-NM	TDJ2E	1', estimation 6"	$\leqslant \pm 6''$	Mag A	0.1 nT	±1~5nT
CTM-DI	CJ6	1', estimation 6"	$\leqslant \pm 4''$	—	0.1 nT	±1nT
GEO-DI	J6	1', estimation 6"	$\leqslant \pm 6''$	—	0.1 nT	±1nT

Based on the aforementioned methodology, the GNC has completed 12 comparisons of DIMs from 2010 to 2024 (no comparison was conducted in 2011, 2013, and 2021), and corresponding datasets have been accumulated. The datasets for comparing all testing instruments with reference instrument are displayed in Fig.2, where Fig.2(a) and Fig.2(b) respectively illustrate the instrumental differences in declination (ΔU_D) and inclination (ΔU_I). Colored dots in the figures represent measurement results from different years, with vertical coordinates indicating instrumental differences, while the dot sizes indicate the standard deviations of measurements, scaled according to the legend on the right. This graphical representation enables a comprehensive evaluation of observational data quality at both individual instrument and the network levels.

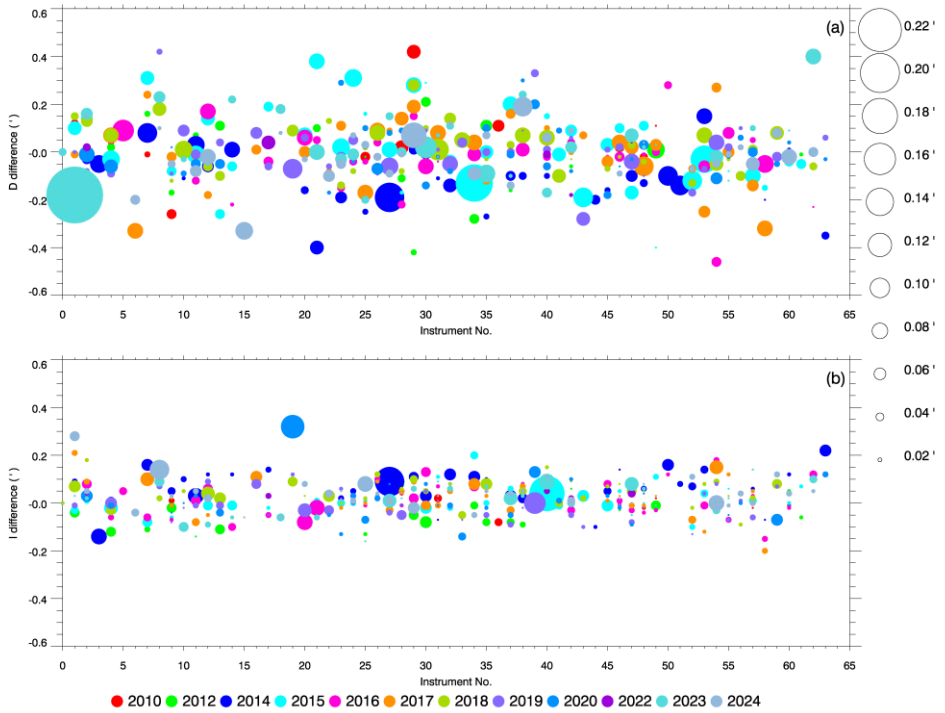


Figure 2: Instrumental differences in declination D (a) and inclination I (b). The dots represent the instrument differences; the size of the dots is the standard deviation with the scale on the right. The colors indicate different years.

This figures also provide insights into instrument performance and operator proficiency. Small dots with large central values indicate significant instrumental differences, suggesting potential instrument malfunctions or operational issues by personnel. Particularly when the difference of D is relatively large while that of I is small, the most likely cause is the positioning error of the theodolite on the pillar, which also reflects 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 accuracy) or inconsistent operational practices. More detailed explanations for tracking abnormal information can be found in He et al. (2009b). This graphical approach thus effectively monitors instrument performance and evaluates observational quality across operators.

To assess of network wide data quality, we conducted statistical analyses of all instrument differences, as shown in Fig. 3. The statistical result 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 a cumulative probability of 90% as the evaluation criteria for the entire network, the corresponding instrument differences thresholds are 0.21′ for D and 0.11′ for I, indicating excellent consistency among network fluxgate instruments. Notably, declination measurements exhibit greater dispersion than inclination values. This difference stems from the additional azimuth marker alignment required in declination

measurements—a process more susceptible to operator error compared to inclination measurements. Another significant source of possible error in declination readings, which is not present in inclination readings, is the accuracy of setting (at 90 ° or 270 °) on the vertical circle.

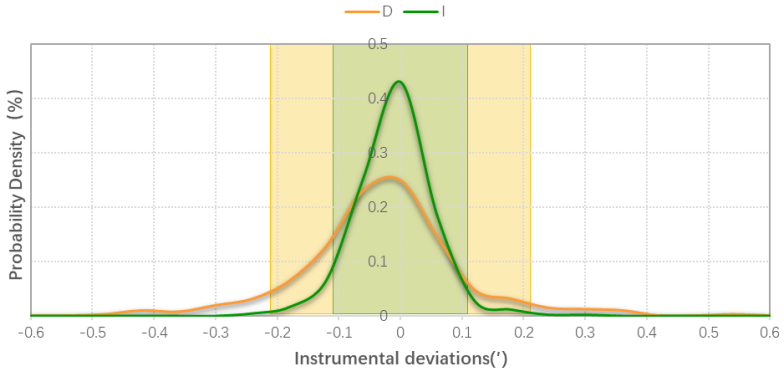


Figure 3: The instrument differences of declination D (orange line) and inclination I (green line).

The dispersion of multiple dots corresponding to the same instrument also reflects its data quality and operation stability. Frequent personnel changes for same instrument have introduced operator dependency errors, manifesting as increased dispersion. To further explore the relationship between frequent personnel changes and the dispersion of instrument differences, the frequency personnel change was defined as the ratio of non-repeated operators to the total number of comparison measurements for each instrument (from Table 1), serving as the x-axis, the dispersion degree was represented by the standard deviation of all instrumental differences for each instrument, serving as the y-axis. To enhance statistical significance, only instruments that participated in 3 or more comparisons were included in the analysis. As shown in Fig. 4(a), the dispersion degree of the instrumental differences increases with the frequent personnel changes, while this phenomenon is less pronounced in Fig. 4(b). This indicates that frequent personnel changes increase observational errors and that personnel changes have a greater impact on D than on I. This result is consistent with the conclusion in the previous paragraph that D errors are larger than I errors. It further provides strong evidence supporting the explanation that operator errors primarily arise from the alignment of markers and the level adjustment of the theodolite.

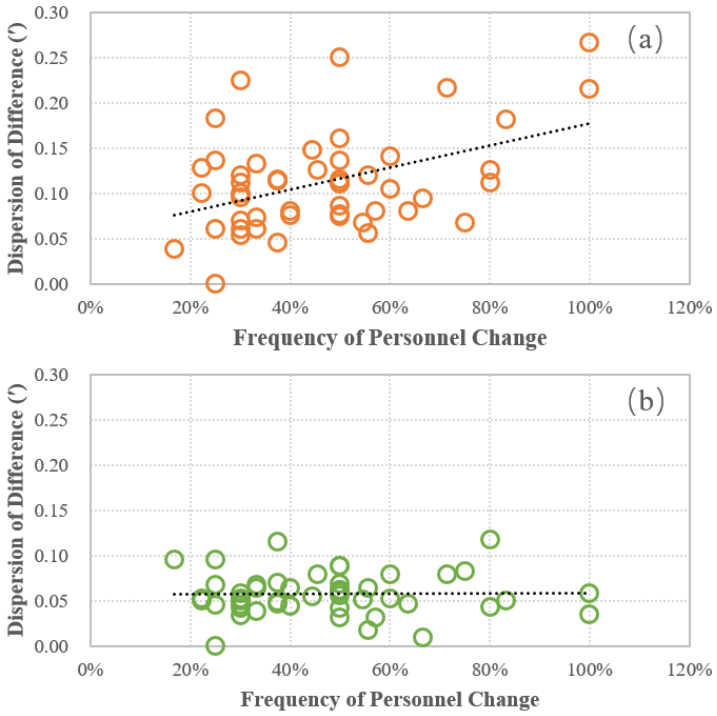


Figure 4: The relationship between frequent personnel changes and the dispersion of instrument differences, (a) declination D, (b) inclination I, and the black dashed line is a linear fitting line.

The instrument differences of all 12 years' comparisons were classified to five group based on the instrument type, and the standard deviations were calculated for each group. This was done to simply compare the stability of observational results across different instrument types, as shown in Fig. 5. It can be clearly seen that MINGEO has better stability, followed by Mag-01H and TDJ2E-NM, while the other two have relatively large dispersion, which is directly related to the resolution of the theodolite.

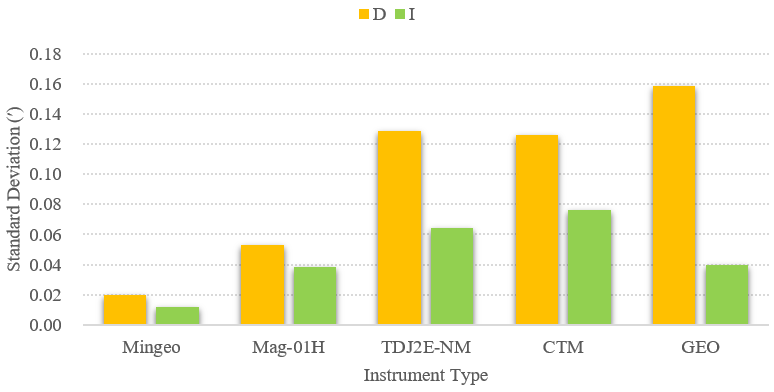


Figure 5: The standard deviation of the instrument differences for five types of instruments.

This simplified comparison measurement provides an efficient mechanism for identifying inter station differences, 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 instrument, it is necessary to analyses long term stability and reliability. The following section will evaluate the reference instruments using all comparison measurement data through uncertainty analysis.

4 Uncertainty analysis on comparison datasets

During the comparison process, fluxgate theodolites are employed to observe magnetic declination and inclination. To minimize errors, each observed magnetic declination and inclination is the results obtained after four measurement processes. And the final instrument differences are obtained by comparing the baseline values. However, errors (as mentioned in section 2.1) cannot be completely eliminated and will still exist, which is the main reason for the instrument differences and the source of uncertainty in measurement results. Therefore, the instrument differences defined in this paper are the comprehensive differences of the entire instrument system, representing the differences between results obtained by the instruments after four measurement processes under the assumption of no personnel operation error. The impact of various internal errors of the fluxgate theodolite on the measurement results is already included in the measurement results and is part of the differences of the theodolites, which will not be discussed separately. Then, a quantitative analysis of the error will be conducted based on uncertainty analysis methods.

As described above, the uncertainty analysis in this process encompasses several key error sources, including internal errors of the theodolite, repeatability errors, operator dependency errors (differences between individuals), and pillar correction errors. Environmental interference is excluded from consideration due to controlled laboratory conditions. The uncertainty analysis process and calculation formula applied in the comparison datasets are described in Appendix B.

For an instrument comparison, after calculating the synthesized internal uncertainty ($u_{inst,i}$) of each instrument, the standard deviation ($s_{rep,i}$) and associated Type A uncertainty ($u_{rep,i}$) of the repeatability error, as well as the standard deviation ($s_{between}$) of all operator-instrument-combinations results, the operator uncertainty (u_{oper}) can be calculated using Eq. (B7) based on these results. When the above calculation process is repeatedly applied to the results of 12 comparison datasets, 12 operator errors are obtained, as illustrated in Fig. 6. The light orange and light green filled areas represent the differences for D and I, respectively. Results show consistently higher operator dependency errors in declination measurements compared to inclination, with D errors persistently exceeding I values. This difference arises from the additional azimuth marker alignment step, and the accuracy of the vertical circle setting (at 90 ° or 270 °) required for declination measurements, which introduces greater operator variability. The mean operator dependency 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 operator dependency errors.

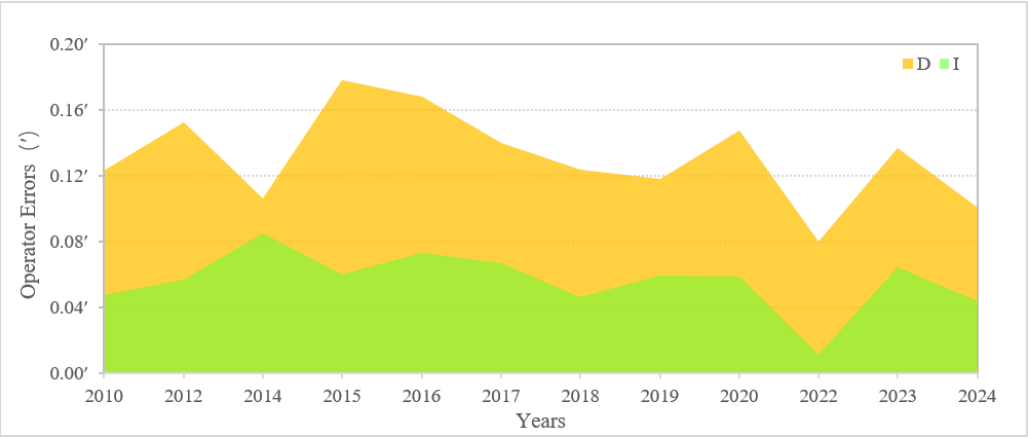


Figure 6: Operator dependency errors in Declination (D) and Inclination (I) measurements.

The robustness evaluation of the reference instrument requires quantification of its systematic deviation relative to the true values and associated uncertainties. If each instrument comparison is regarded as an independent measurement experiment, then the average of the measurement results (after correction) from all instruments (excluding the reference instrument) can be taken as the true value of that measurement experiment. Then, the measurement result from the reference instrument can be compared with the true value, and the total uncertainty of the comparison result can be calculated based on the uncertainty formula. In this way, the operational status of reference instruments can be examined. By applying this same process to 12 sets of comparison data, the long-term stability of the reference instrument can be evaluated.

Using the methodology described above, we analyzed 12 years of instrumental difference data for declination (D) and inclination (I). The time series of mean differences (Δ) between the reference instrument and all instruments, along with their uncertainties (u_Δ), are shown in Fig. 7. In this 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 $|\overline{\Delta}| \leq 2u_{\overline{\Delta}}$, most mean differences fall within the $2u_\Delta$ range.

However, mean differences for 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. (B16) and (B17), 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 $|\overline{\Delta}| \leq 2u_{\overline{\Delta}}$ (95% confidence level), both D and I meet this requirement, confirming the reliability of the reference instrument's long term observational data.

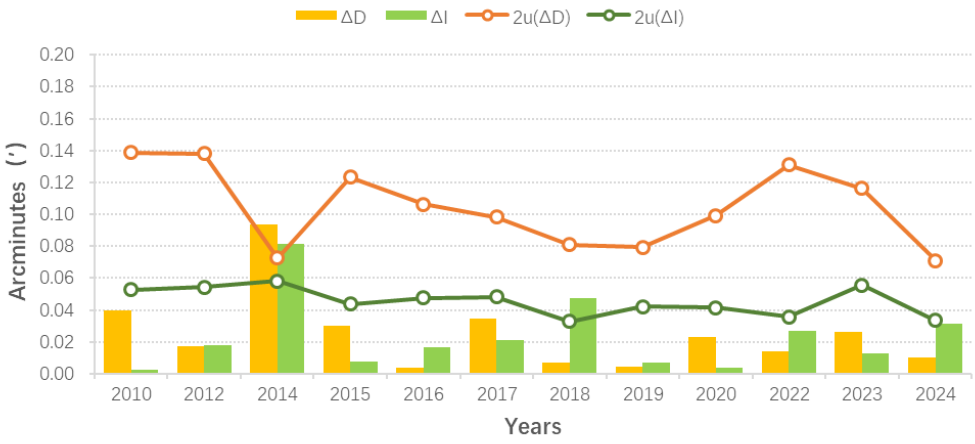


Figure 7: Time Series of Mean Differences and Uncertainties Between Reference and Tested Instruments.

5 Conclusion and outlook

This study systematically analyzed 12 sets of comparison measurement data from the GNC. The research findings demonstrate the critical importance of instrument comparisons in geomagnetic observatory networks, primarily manifested in the following two aspects: (i) The instrument comparison effectively monitors the operational status of instruments at various observatories, assesses the technical proficiency of operators, and enables timely repair and maintenance of problematic instruments; and (ii) The comparison data facilitates the analysis of factors contributing to absolute geomagnetic observation errors, allows for a comprehensive evaluation of data quality across the entire geomagnetic network, and provides a means to assess whether reference instrument meets standards. These assessments also offer valuable references for evaluating the reliability of scientific conclusions derived from geomagnetic data.

The statistical analysis results reveal that when the probability density of instrument differences accumulates to 90%, the corresponding instrument difference are 0.21′ (D component) and 0.11′ (I component), which can serve as evaluation criteria at the network level. The statistical results of the comparison also effectively reflect the factors influencing the declination measurement. Specifically, they reveal that the frequent rotation of operators has a significant impact on the declination observation results, while its effect on the inclination is not as pronounced. This further suggests that operator dependency error is the primary source of error in absolute geomagnetic measurements.

Through uncertainty analysis of multi-source error, the systematic differences between the reference instruments and the test instruments were quantified. The operator dependency errors of D (0.13′) and I (0.06′) were successfully separated and

consistent with the observed experimental results, confirming that operator dependency error is the primary factor contributing to instrument differences. Notably, operator dependent errors in D are significantly higher than I, which once again proves that D is more susceptible to human factors. A comprehensive evaluation was developed to assess the robustness of reference instrument. By constructing time series of mean differences (Δ) and their uncertainties between the reference instrument and all testing instruments, the long term stability of the reference instrument can be analyzed. The results showed that the data remained stable for a long time without significant drift, and the average of mean differences are -0.004′ (D) and 0.022′ (I), both within the 95% confidence interval. This indicate the reference instrument exhibit high stability and reliability.

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 comparison 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.

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 dependency 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) in the future, it is anticipated that operator dependency differences will be eliminated, thereby obtaining higher quality geomagnetic observational data.

Appendix A: Measurement procedure and comparison methodology

A.1 Measurement procedure

Geomagnetic declination and inclination measurements are performed within the horizontal and magnetic meridional 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. In order to eliminate errors associated with the optical misalignment of the theodolite, two observations are required to find the true north direction, one with sensor up and the other with sensor down. Finally, the direction of the azimuth marker can be determined through two readings and recorded as M (Fig.A1). As the azimuth value A of the marker known, the true north position can be calculated. Subsequently, the geomagnetic meridian direction is identified by searching the fluxgate null position in the horizontal plane (with the vertical circle maintained at 90 °or 270 °) and recording

horizontal reading (D'). The geomagnetic declination D is then derived from the differential angular measurement following the formula:

$$D = D' - M + A \ . \quad (A1)$$

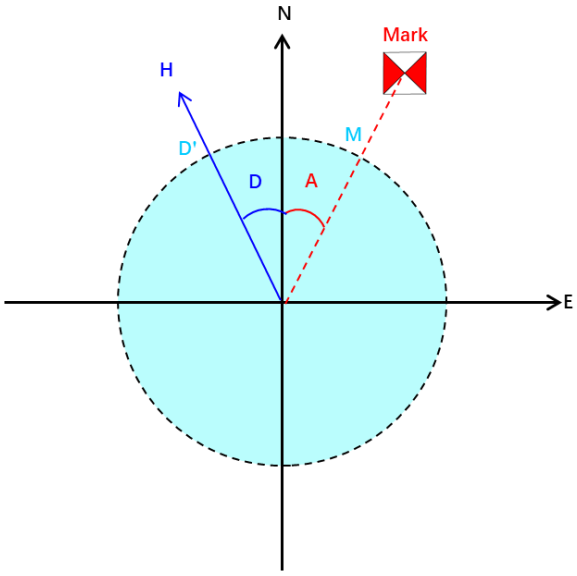


Figure A1: Measurement principle of the declination.

Inclination measurements follows analogous procedures and is carried out in the magnetic meridional plane derived from the previous declination measurements, while also within the vertical reference provided by the gravity field through the theodolite suspension system.

The measurement procedure follows the guide published by IAGA (Newitt et al., 1996), and the specific description of the four positions observation can refer to the observation steps in Csontos and Sugar's (2024) paper. The declination measurement protocol is preceded and followed by sensor up and down azimuth marker readings and then involves four configurations: (i) telescope 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). Four different position observations can eliminate errors associated with theodolite optics, sensor misalignment and electronics offset (Csontos and Sugar, 2024). Then final declination value is derived through arithmetic averaging:

$$D' = (D_1 + D_2 + D_3 + D_4)/4 + 90^\circ \quad . \quad (A2)$$

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:

$$I = (I_1 + I_2 - I_3 - I_4)/4 + 90^\circ \quad . \quad (A3)$$

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.

By using the geomagnetic declination (D), inclination (I), and the total magnetic intensity (F) measured by the proton magnetometer, all the absolute components of the Earth's magnetic field can be calculated. This will facilitate the subsequent baseline calculations of variometer for all components (such as east, north, and vertical directions).

A.2 Baseline calculation and comparison

The formula for calculating the baseline value of component W, as defined in the INTERMAGNET Reference Manual (St Louis, 2024), is presented below:

$$W_B(k) = W_o(i{:}j) - W_R(k), \quad \text{(A4)}$$

where $(i{:}j)$ is the time interval (typically minutes) for measurement, (k) is the k-th time, the average time of interval $(i{:}j)$, $W_o(i{:}j)$ is the absolute field value for the time interval $(i{:}j)$, $W_R(k)$ is the variometer recorded value at time k, and $W_B(k)$ is the derived baseline value.

When absolute measurements are performed on different pillars, baseline correction to the reference pillar requires pillar differences. The generalized formulation for component W correction is:

$$\Delta U_{SO} = W_{BS} - W_{BO} + \Delta W_{SO}, \quad \text{(A5)}$$

where s is the reference pillar designation, o is the non-reference observation pillar, W_{BS} and W_{BO} are respectively the baseline values from reference and observation pillars, ΔU_{SO} is the final instrument difference, ΔW_{SO} is the pillar difference, which represents the difference between the base pillar and the reference pillar. These pillar differences were measured before the observatory was put into operation and remeasured before each comparison.

There are two main ways to determine the pillar difference: the direct simultaneous measurements and indirect baseline values comparison. If two or more instruments are available for measurements, the direct method can be applied, and the pillar difference can be calculated by the following Eq. (A6):

$$\Delta W_{SO} = \left[\left(W_{ps} + W_{qs}\right) - \left(W_{po} + W_{qo}\right)\right]/2 \ , \quad \text{(A6)}$$

where p, q denote the different instruments and s, o represent the standard pillar and other observation pillars, respectively. W_{ps} and W_{qo} represent the baseline value of instrument p on standard pillar and the baseline value of instrument q on other pillars, respectively. ΔW_{SO} is the pillar difference between the standard pillar and other pillars.

If only one instrument is available, the indirect method can be used to calculated the pillar difference using the follow Eq. (A7):

$$\Delta W_{SO} = W_{ps} - W_{po} \quad . \quad \text{(A7)}$$

This methodology enables cross comparisons of fluxgate theodolite through pillar reference baseline correction. The obtained difference values ΔU_{SO} provide quantitative evaluation parameters for assessing absolute observation data quality across participating instruments.

Appendix B: Application of uncertainty in comparison datasets

The true value is the absolute, unbiased value of the measured physical quantity, which is typically unattainable directly in most cases. It can usually be represented by the arithmetic mean of sufficiently repeated measurement data, while uncertainty characterizes the dispersion of the measured value, indicating the range within which the true value may lie, including Type A and Type B standard uncertainties. Type A uncertainty is a type of uncertainty evaluated through statistical methods (e.g., standard deviation of repeated measurement data) to assess the reliability and dispersion of measurement results. Its evaluation relies on the statistical analysis of repeated experimental data. While Type B uncertainty is based on non-statistical methods (e.g., instrument calibration certificates, empirical formulas, or known error limits), often combined with prior information or professional judgment (ISO/IEC GUIDE 98-3:2008).

B.1 Uncertainty of fluxgate theodolite error

The fluxgate theodolite consists of two primary components: the theodolite and the fluxgate sensor. Therefore, its uncertainty of internal error also need to be calculated separately for each component. For the theodolite component, errors usually follow a normal distribution, while for the fluxgate sensors, errors raised from the limited resolution (ε=0.1 nT) of the liquid crystal display follow a uniform distribution. Both can be evaluated using the Type B standard uncertainty according to ISO/IEC GUIDE 98-3:2008. Typically, the maximum permissible standard deviation of horizontal and vertical angles for one measurement cycle is considered as the standard for theodolite level classification based on GB/T 3161-2015 (Standardization Administration of China, 2015). This can be used as a parameter to evaluate the theodolite. Therefore, the first step is to consult the theodolite's relevant manual and obtain its parameter. Based on the data in Table 2, the Type B standard uncertainty for theodolite component and fluxgate sensor can be calculated using Eq. (B1) and Eq.(B2) from ISO/IEC GUIDE 98-3:2008., respectively.

$$u_{b1} = \delta/3, \quad (B1)$$

where δ is the maximum permissible standard deviation.

$$u_{b2} = \varepsilon/\sqrt{3}, \quad (B2)$$

where ε is the limited resolution.

In addition, there may be other uncertainties that affect the observations, such as magnetic contamination of the theodolite body. Although these effects are difficult to quantify, they increase the uncertainty of the measurement results and will also be reflected in the measurement results. So the synthesized internal uncertainty for each instrument is computed as the root sum square of two elements, as shown in Eq. (B3):

$$u_{inst,i} = \sqrt{u_{b1,i}^2 + u_{b2,i}^2} \quad . \tag{B3}$$

B.2 Uncertainty of repeatability and operator error

Operators from different observatories performed repeated observations (i.e., 6~8 sets of results) using their own instruments during each instrument comparison, thereby obtaining results of the operator-instrument combination. Based on these results, the standard deviation ($s_{rep,i}$) and associated Type A uncertainty ($u_{rep,i}$) of the repeatability error for each operator-instrument-combination can be calculated using Eq.(B4):

$$u_{rep,i} = \frac{s_{rep,i}}{\sqrt{N}}, \tag{B4}$$

$$s_{rep,i} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(x_{i,k} - \frac{1}{N}\sum_{k=1}^{N} x_{i,k}\right)^2}, \tag{B5}$$

where x is the baseline value calculated according to Eq. (A4) and corrected for pillar difference, N is the number of baseline values for each instrument.

Since multiple operator-instrument-combinations from different observatories are involved in an instrument comparison, the standard deviation ($s_{between}$) of all operator-instrument-combinations results can be calculated following Eq. (B6). It includes both the internal error of theodolite and the operator error, so the operator uncertainty (u_{oper}) can be derived by subtracting the averaged instrumental uncertainties, as shown in Eq. (B7):

$$s_{between} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\overline{x_i} - \bar{\bar{x}})^2} \ , \qquad \bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\overline{x_i} \ , \tag{B6}$$

$$u_{oper} = \sqrt{s_{between}^2 - \frac{1}{N}\sum_{i=1}^{N} u_{inst,i}^2} \ , \tag{B7}$$

where $\overline{x}$ is the average baseline value of each instrument, N is the number of instruments involved in the comparison.

B.3 Uncertainty of pillar correction error

The geomagnetic absolute observation room usually has several observation pillars, one of which is the standard pillar (pillar No. 1# in this paper). Although the magnetic field gradient in the observation room is very small, there are still differences in the magnetic field between different observation pillars, which called pillar difference. This means that the data observed on other pillars need to be converted to the standard pillar through pillar difference correction before being compared with the data observed on standard pillar. So the uncertainty of pillar correction error must be considered in the uncertainty analysis.

In comparison, different instruments may be installed on different observation pillars. To unify the observation results of these instruments to the standard pillar for comparison with the standard instrument's results, pillar difference corrections must be applied to the measurement data of each instrument. These pillar differences (ΔW_{SO}) and their uncertainty (u_{pier}) can be obtained from the measurement results of pillar differences at each observatories. They are obtained by using repeated measurement data and calculating according to Eq. (A6) or Eq. (A7). They can also be checked using all the comparison data.

Given that the magnetic gradient within the observation room is very small, the pillar difference is therefore typically minimal. Nevertheless, prior to initiating each comparison process, it was remeasured to ensure accuracy. Table B1 presents the pillar differences and their uncertainties at the observatory where the comparison is conduced.

Table B1: Pillar differences and their uncertainties

Pillar No.	Pillar Difference (ΔW_{SO})		Uncertainty (u_{pier})	
	D(′)	I(′)	D(′)	I(′)
1#	0.00	0.00	0.01	0.01
2#	0.06	-0.03	0.05	0.02
3#	0.10	0.15	0.04	0.03
4#	0.17	0.07	0.03	0.02
5#	0.13	0.28	0.07	0.04
6#	0.30	0.11	0.03	0.02

B.4 The total synthesized uncertainty

The total synthesized uncertainty for each instrument is then aggregated as the root sum square of all contributing factors, according to Eq. (B8):

$$u_i = \sqrt{u_{inst,i}^2 + u_{oper}^2 + u_{rep,i}^2 + u_{pier,i}^2} \quad . \tag{B8}$$

Finally, the ensemble mean (μ_{group}) and combined uncertainty (u_{group}) for all instruments are computed using a weighted average approach. Weights (ω_i) are assigned inversely proportional to the square of each instrument's total uncertainty, ensuring higher precision instruments exert greater influence, using Eq. (B9) and (B10):

$$\mu_{group} = \frac{\sum_{i=1}^{N} \omega_i \mu_i}{\sum_i^N \omega_i} \quad , \quad \omega_i = \frac{1}{u_i^2} \quad , \tag{B9}$$

$$u_{group} = \frac{1}{\sqrt{\sum_i^N \omega_i}} \quad . \tag{B10}$$

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

B.5 Multi years comparison analysis method

Building on the uncertainty analysis of single comparison sessions, this section evaluates the long term stability and robustness of the reference instrument using data accumulated over 12 comparisons within the GNC. Each comparison 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 (μ_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. (B11) and (B12):

$$u_{rep,s} = \frac{s_{rep,s}}{\sqrt{N}} \quad , \tag{B11}$$

$$s_{rep,s} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(x_{s,k} - \mu_s\right)^2} \quad , \quad \mu_s = \frac{1}{N}\sum_{k=1}^{N} x_{s,k} \quad . \tag{B12}$$

The total uncertainty of the reference instrument (u_s) incorporates its internal error ($u_{inst,s}$), the operator dependency uncertainty (u_{oper}), and repeatability uncertainty, as expressed in Eq. (B13):

$$u_s = \sqrt{u_{inst,s}^2 + u_{oper}^2 + u_{rep,s}^2} \quad . \tag{B13}$$

The difference (Δ) between the reference instrument and the weighted ensemble mean (μ_{group}) of all participating instruments, along with its uncertainty (u_Δ), is quantified for each session using Eq. (B14) and (B15):

$$\Delta = \mu_{group} - \mu_s \quad , \tag{B14}$$

$$u_\Delta = \sqrt{u_s^2 + u_{group}^2} \quad . \tag{B15}$$

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

To assess long term stability, differences (Δ) from M years comparison sessions are compiled into a time series plot, enabling visual detection of potential drifts caused by environmental fluctuations or instrumental aging. The multi-year mean difference ($\overline{\Delta}$) and its uncertainty ($u_{\overline{\Delta}}$) are calculated through Eq. (B16) and (B17):

$$\overline{\Delta} = \frac{1}{M}\sum_{m=1}^{M}\Delta_m \quad , \tag{B16}$$

$$u_{\overline{\Delta}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M} u_{\Delta,m}^2 + \left(\frac{s_\Delta}{\sqrt{M}}\right)^2} \quad , \tag{B17}$$

where, (s_Δ), 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.

Data availability

The raw data are available upon request from the corresponding author at zxd9801@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.

Csontos, A. A., and Sugar D.: Dataset of geomagnetic absolute measurements performed by Declination and Inclination Magnetometer (DIM) and nuclear magnetometer during the joint Croatian-Hungarian repeat station campaign in Adriatic region, Data in Brief, 54:110276, doi.org/10.1016/j.dib.2024.110276, 2024.

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, 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.

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.

International Organization for Standardization and International Electrotechnical Commission: Uncertainty of measurement, Part 3: Guide to the expression of uncertainty in measurement (GUM:1995), ISO/IEC GUIDE 98-3:2008, https://www.iso.org/resources/publicly-available-resources.html (last access: 2 September 2025), 2008.

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

Lauridsen K.E., in: Experiences With the Dl-fluxgate Magnetometer inclusive theory of the instrument and Comparison With Other methods, Geophysical Papers R-71, Danish Meteorological Institute, Copenhagen, p. 30, 1985.

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.

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.

Okada, M., Toya, T., Koike, 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.

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.

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.