Comments on "Evaluating data quality and reference instrument robustness: insights from 12 years DI magnetometer comparisons in the Geomagnetic Network of China" by Y. He *et al.*

The authors have improved the manuscript in response to the points raised by the two reviewers. I find the contents of the revision themselves are sufficient. As the reviewer noted, what makes this paper highly unique is their valuable dataset acquired over a long period at numerous observatories across China. Potentially providing useful information to the geomagnetic observation community, the paper itself is considered suitable for publication.

To make the paper publishable eventually, please address the following two points. Given the scale of the required changes, I impose them as a major revision. While this will place additional burden on the authors, I expect they will make a positive effort to achieve publication.

First, it is unclear what the term "reference instrument" which is the subject of this paper and included in the title, refers to. It first appears in line 51 of Section 1, but it is only a "reference instrument" in a general sense, and can even be read as referring to a proton magnetometer. Section 1 should state that DIM No.0 (the "standard reference instrument" mentioned in line 171) is evaluated as the example for this study.

Table 1 shows "GNC" as the observation station code for DIM No.0. It is unclear at which specific station's reference pillar the measurements were made. Does this mean that the tested DIMs were brought to a specific central station for comparative observations? If so, please clearly highlight the central station in Figure 2. Since pillar differences across remote observatories are generally unobtainable, I presume that the tested DIMs were gathered at one location each year for the intercomparison. If the inter-station DIM comparison were made remotely, clearly describe so in the subsection 2.4, as well as in the abstract.

Second, on top of the original inadequacy of the paper's structure, and the substantial additions made during revising process greatly increased its length, making the author's arguments very difficult to follow (actually, there are so many "Additionally" and "Furthermore" throughout). The presentation should be refined by addressing the following points:

- Rewrite the abstract which is currently seriously disorganized. Just give the outline of paper concisely. No specific figures are necessary. No outlook is necessary. Get straight to the reference instrument robustness. The text of the conclusion section rather has the taste of an abstract. Is the period 12 years or 15 years?

- Keep Section 2 dedicated only for the methodology, as its title says. The three subsections 2.1 to 2.3 review the absolute measurement. I don't think they are essential for discussing the author's findings. They could be eliminated, or compressed to form a subsection in Appendix.

- Let Subsection 2.5 be independent from Section 2 for the methodology. It can be presented in a Results section.

- The authors could move the details of Sebsections 3.1 and 3.2 to Appendix, describing the specific formulations for the uncertainty estimation as an application of the ISO guideline for the current analysis. In the main text, the procedure can be introduced simply, referring to Appendix. It would be the subsection 3.3 that the authors would like to highlight the most and draw reader's attention to.

The above are my suggestions, which are not the condition for publication. The authors do not have to take all of them as they are. In those cases, nevertheless, replies with reasonable explanations are expected.

Minor issues

Line 246: A new paragraph would start with "This study..."

Line 288: Incomplete sentence.

Line 288: Misspelling "conduced".

Table 2: Misspelling "Resulution".

Line 467: The authors' first names and last names are reversed.