

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2022-203-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-203**

Anonymous Referee #2

Referee comment on "Rescue and quality control of historical geomagnetic measurement at Sheshan observatory, China" by Suqin Zhang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-203-RC2>, 2022

Suqing Zhang and coauthors have submitted a manuscript entitled 'Rescue and quality control of historical geomagnetic measurement at SheShan Observatory, China' to ESSD. The manuscript deals with the digitization of hourly mean values of SheShan Geomagnetic Observatory from 1933 to 2019.

The paper is a valuable contribution to the global geomagnetic observatory data set and the authors are to be commended for their efforts. The analysis presented in this manuscript is able to show some general data quality features and confirms to a certain degree that the data quality is good. The method of this analysis is sometimes of an ad hoc approach (like the outlier identification in Excel) and the authors would profit in future studies if they use more scientifically and geomagnetically motivated methods like comparison of the H component with the Dst index and the use of the time derivatives of the hourly means. I recommend minor revisions for the present manuscript as indicated below.

The abstract is well written and the English reads very well. Still, for publication its language needs improvement. Even in the abstract, there are a number of mistakes including missing and superfluous blanks. I will not mark or correct language mistakes in detail here, as they are too many. But before publication, it is the duty of the authors to read again the complete article and correct the language. I am happy to read the article once more after that.

I have not heard about zenedo before, but it looks like it is supported by Cern and is a trustworthy data archive and distribution instance.

You provide, through zenedo, a 50 MB excel spreadsheet. My Office package crashed when I tried to load this file. My estimate is that if you provide HMVs (hourly mean values) in a simple ASCII format, then 100 years of data will be only about 10 MB in file size.

Excel seems to be not an ideal format. In any case, smaller yearly files would be more appropriate.

Line 56

What is the meaning of 'correction of the selected homogeneity'?

64

What is 'reference room'?

69

What is the DBF format?

Same for BAS and MDB, please explain in the text.

102

Please rewrite 'useful for re-verifying the data when the documents were not available', it took me along time to understand what you want to say. Also, I guess that for old paper copies it is not good to be carried around too much and as soon as you have a digital picture, which is fast to make, you can bring the respective paper again to its normal archive place with the usual temperature, humidity etc.

Also, which character recognition program did you use?

I agree that typing the numbers by humans is often a good solution.

111

What is 'large value'?

I think what you mean is a 'base value', to be added to the tabulated values to get the AHMVs. You need to explain what the base value is as most people do not know the concept.

115

I do not understand:

'After half a year's efforts, we input the data from 1933 to 1954 into the Excel template and

completed the digitization of paper data.'

Do you mean:

'Using this approach, it took us half a year to digitize the 1933 to 1954 data from paper records.

125

I would move the reference to Zhang 2016 behind 'Geomagnetic Network of China', to give the reader a reference for the Geomagnetic Network of China.

137

I agree that a very important property of geomagnetic observatory data (and geophysical time series in general) is to be homogeneous. But I disagree with the use of the term 'inhomogeneities' in this context, even though it was used by Morozova et al (2014), too.

For example, if you have a time series of 10 years and the first 5 years it was recorded by an instrument A and the last 5 years by an instrument B and their transfer functions are different. I agree that the time series is not homogeneous, but I do not know what would be the homogeneous part of the time series and what part is the inhomogeneity.

Suggestion: call it data problems and define it as 'sudden breaks and jumps in the series of geomagnetic data, or gradual biases, or noise and change of transfer function etc.'

142

continuity -> completeness, or data availability

147

having a total of 67 monthly data missing -> having a total of 67 months of data missing

148

Remove '. "Every trial to correct data can produce unwished secondary effects in the result " (Linthe, 2013),' if you do not have data, you cannot correct data. Data interpolation is not permissible for AHMVs as the temporal change in the magnetic field is happening on periods longer than 1 hour and shorter than the length of your data gaps.

163

The noise -> The additional signal

209

AP -> Ap

page 13

The FTD of AHMVs should have been analyzed instead of FTD og ADMVs, as the daily means are much more smoothed than the hourly means.

247

corrected -> correction

page 14

A jump in the data cannot be corrected by determining the difference of the mean 3 month prior and 3 months after the jump. With this approach, you include errors from secular variation, from seasonal variation and from geomagnetic disturbance. You are not removing the jump-problem, you just make it less visible by subtracting a value close to the unknown jump. I recommend to not correct the data here, or to determine the exact jump from the measurements during the change of instruments

264

Jump outliers are best detected by time derivative of AHMVs, not by comparison with models.

299

and so on -> and data quality problems in SSH.

322

You can not see a potential jerk in 2011 because of your data gap until 2011.

330

You can also include the study of geomagnetic storms.