Supporting Information for "Seven decades of neutron monitors (1951 – 2019): Overview and evaluation of data sources"

Pauli Väisänen[1], Ilya Usoskin[1,2], Kalevi Mursula[1]

[1]Space Physics and Astronomy Research Unit, University of Oulu, Finland

[2]Sodankylä Geophysical Observatory, University of Oulu, Finland

Contents of this file

1. Text S1
2. Figures S2 to S4

Additional Supporting Information (Files uploaded separately)

1. Neutron monitor information table S5 (NMStationInfoList.xlsx)
2. Data source recommendation list S6 (NMSourceRecommendations.xlsx)

Introduction This supporting information contains information and examples about the data analysis process of the main article. The station recommendations and station information which cannot be incorporated in the main article text are also included here. A description for all supplementary information is given.

Text S1. Here we give an overview of the analysis of the quality of NMs on an example of the Newark station, referred henceforth as NWRK.

Figure S2 shows the original data, the normalized data and the ratios between them, corresponding to different sources. One can see from the upper panel that the datasets are almost equivalent in terms of shape, but the absolute level of datasets, retrieved from the NMDB source, is higher than the others (WDCCR line is at the same level as IZMIRAN and the Station's web-site), by a constant factor of 5/3. After the data normalization to the median of years 1995 – 1996 (middle panel), the datasets become nearly identical, except for some differences related to the data coverage after 2017 and a small step-like difference between NMDB and the others in 2012 – 2015. The ratios (low panel) between the normalized datasets are very close to unity during most of the time, except for small discrepancies across the dataset. The offset-type difference in 2012 – 2015 is about 4 %.

Figure S3 depicts ratios between the hourly values of the normalized count rates of the NWRK NM, obtained from the different sources and those of the reference dataset $\mathrm{NM_{med}}$. For WDCCR, the data for Swarthmore (the previous name of the NWRK station) is also shown. First, one can see that the length of the NWRK dataset is different in different sources. It is the longest (since 1964) in WDCCR, IZMIRAN, station's webpage and NMDB *1hr*, but shorter (since 2000) in the NMDB-revori data tables. There are also some outliers (red points in the Figure), defined as hourly values which deviate by $> 10\%$ from the normalized reference $\mathrm{NM'_{med}}$ values. Such outliers are relatively frequent after 2010 in the NMDB-revori table but absent in the WDCCR and the station web-page. However, the latter two do not contain the data since 2017. NMDB and IZMIRAN have data also after 2017 and contain several outliers. Another important aspect is the long-

term stability of the data. While NMDB-1hr dataset is consistently tied to unity, other datasets exhibit some apparent features, which are not outliers: a 4 % drop during 2012–2015 in WDCCR, IZMIRAN and station's tables, or a long-term trends the NMDB-revori table. This illustrates that datasets from different sources are not identical, and need to be carefully checked. For the case of NWRK NM we recommend that the NMDB-1hr dataset is used as the data source. This recommendation is given even in spite of the 5/3 ratio difference, since NM data are mostly used for analyzing relative and not absolute changes, and the difference is very easily corrected if needed.

A similar analysis has been performed for all the 147 stations in the study, and the resulting list of recommendations S5 is based on all of them.

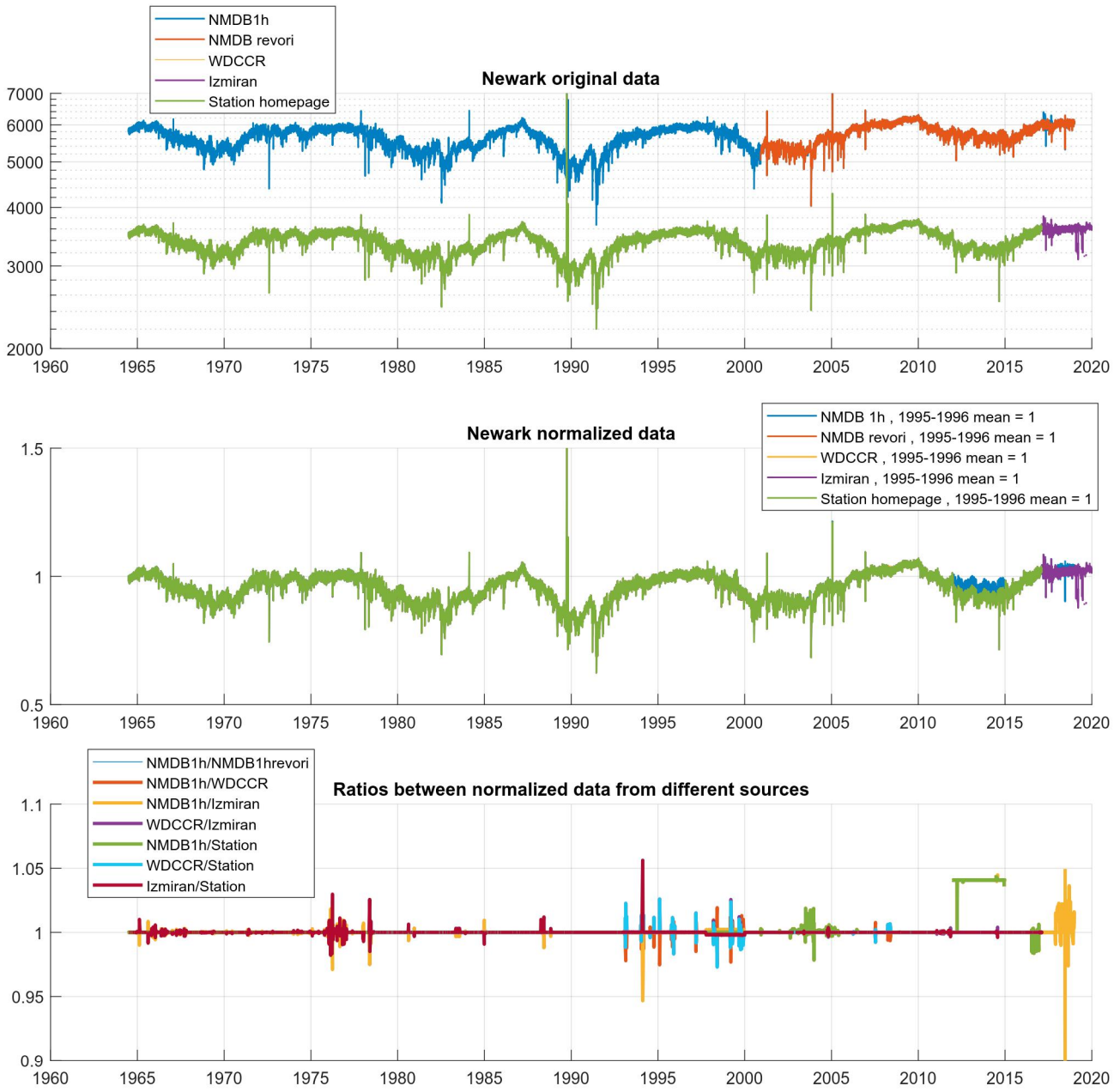

Figure S2. Top: Original Newark/Swarthmore (pressure-corrected) data from different sources. Middle: The same data normalized to the median of years 1995 – 1996. Bottom: Ratios between the normalized datasets, shown in the middle panel.

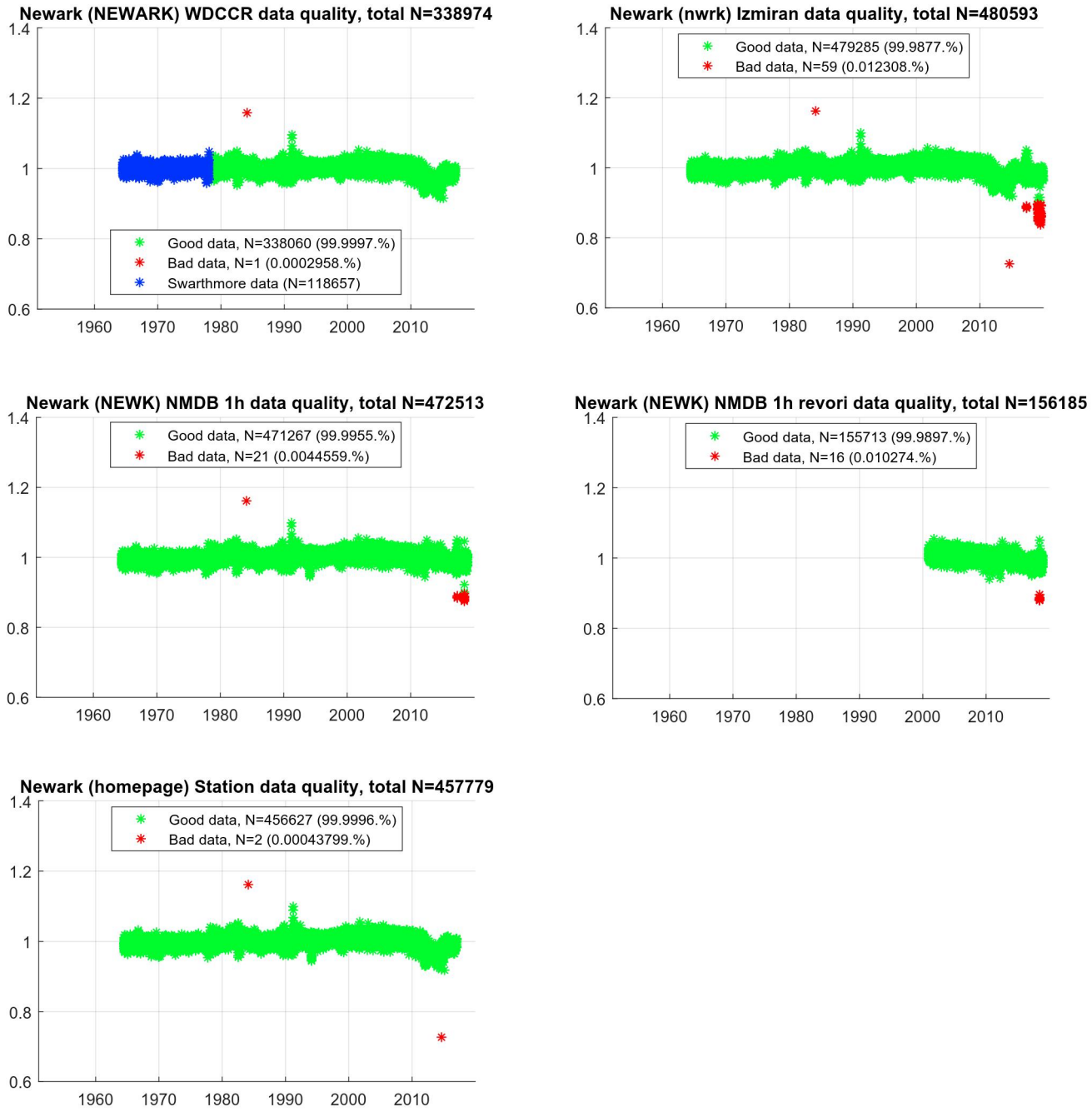

Figure S3. The ratio of the normalized Newark (NWRK) NM hourly count rates, as obtained from different data sources, to those of the reference $\mathrm{NM}'_{\mathrm{med}}$ dataset (Figure S4b). The NWRK data sources are WDCCR, NMDB-1hr, NMDB-revori, IZMIRAN, and station's site, respectively. The color indicates the data quality: *green* indicates good data within $\pm 10\%$ of the $\mathrm{NM}'_{\mathrm{med}}$ values. The total number and the percentage of hourly data points of different quality are given in the legends.

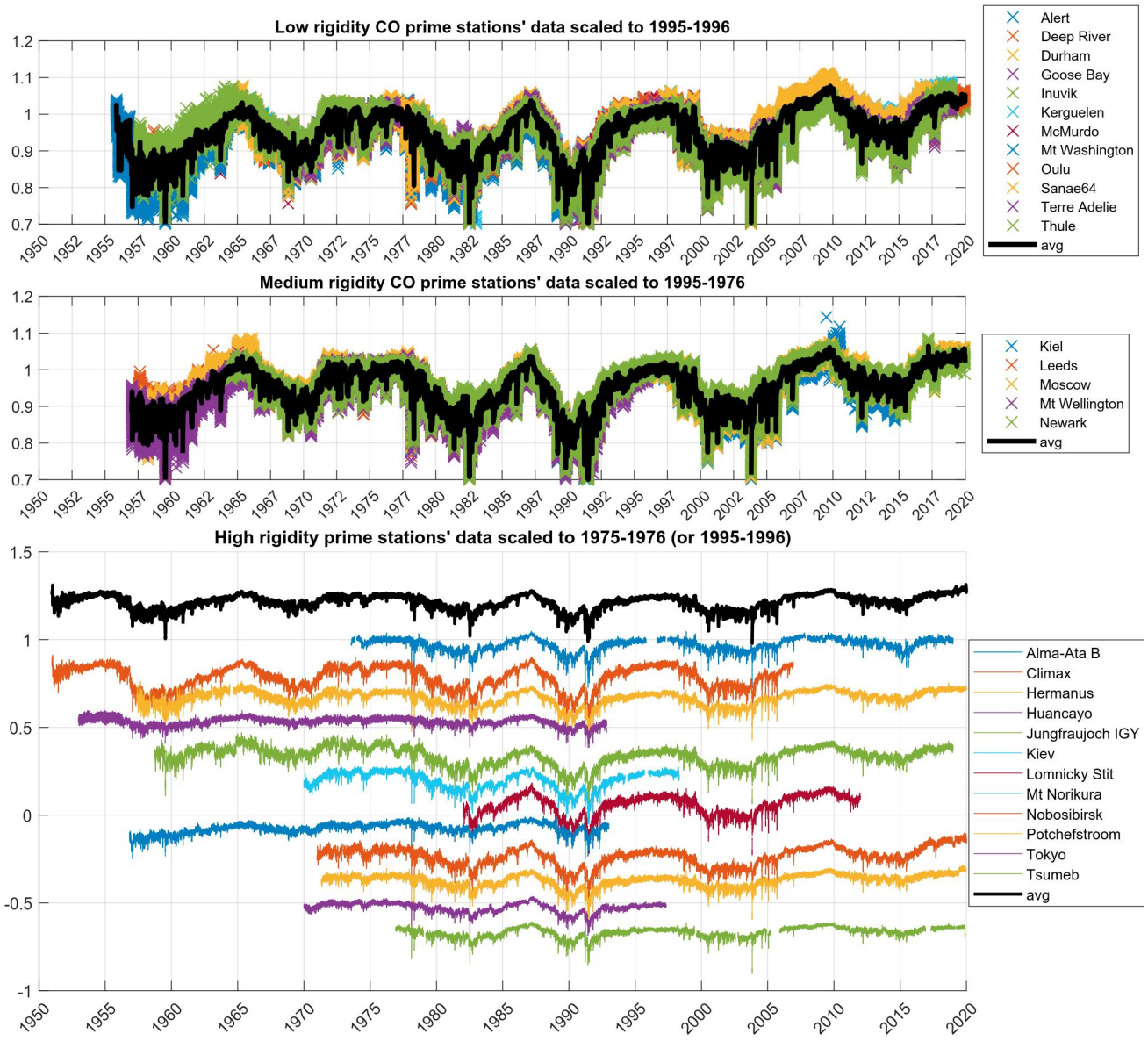

Figure S4. Normalized (see text) time profiles of hourly datasets for the primary NM stations for low ($R_c \leq 1.75$ GV, upper panel) and medium ($1.75 < R_c \leq 2.75$ GV, middle panel) rigidity groups. Colored curves depict individual datasets, while black ones represent the group-averaged reference dataset. High rigidity stations (below) are each offset by 0.15 from each other. Full description is given in the main article Section 4.

Table S5. The data table named NMStationInfoList.xlsx contains metadata, coverage and data quality information. It includes the following 29 columns:

1. Station name	16. Geographical longitude
2. Other name(s)	17. Altitude of the NM location
3. Start year - End year	18. Geomagnetic Cut-off rigidity
4. WDCCR acronym	19. Number of "Good" data points in WDCCR
5. NMDB acronym	20. Fraction of "Good" data points in WDCCR
6. IZMIRAN acronym	21. Number of "Good" data points in NMDB-1h
7. URL of Station homepage	22. Fraction of "Good" data points in NMDB-1h
8. Number of available data sources	23. Number of "Good" data points in NMDB-revori
9. Data points (1h) in WDCCR data	24. Fraction of "Good" data points in NMDB-revori
10. Data points (1h) in NMDB-1h data	25. Number of "Good" data points in IZMIRAN
11. Data points (1h) in NMDB-revori data	26. Fraction of "Good" data points in IZMIRAN
12. Data points (1h) in IZMIRAN data	27. Number of "Good" data points in Station
13. Data points (1h) in Station data	28. Fraction of "Good" data points in Station
14. Maximum coverage in years	29. Maximum coverage by "Good" data in years
15. Geographical latitude	

Table S6. The table named NMSourceRecommendations.xlsx (uploaded separately) contains the recommended and secondary source of neutron monitor data alongside short notes regarding the dataset.