Description of digital values of the real-time AE index

WDC for Geomagnetism, Kyoto
operated by
Data Analysis Center for Geomagnetism and Space Magnetism
Graduate School of Science, Kyoto University

Original version: January 31, 2025
Updated: April 9, 2025

Since December 2024, the digital values of the real-time AE indices have been published in addition to their quick look plots.

Use guidelines

The digital values of the real-time AE index can be used for scientific research in the same way as the provisional AE index. However, real-time AE data from periods when digital values are not published are intended for monitoring, diagnostic, and forecasting purposes only, as they are based on raw data used without visual checks. Please note that users should specify the data version and cite the data DOI when using any version of the AE index. Users should use the highest available version of the index (i.e., the final/provisional AE index, if available). Therefore, please check whether the final/provisional AE index is available before using the real-time AE index.

The digital values have been published on the website `https://wdc.kugi.kyoto-u.ac.jp/ae_realtime/data_dir/` with a delay of up to three weeks. Data for each day are located in `/data_dir/YYYY/MM/DD/` (where YYYY is the 4-digit year, MM the 2-digit month, and DD the 2-digit day). Each data file contains one component of the AE indices (AU, AL, AE, or AO) for a single day, formatted similarly to the WDC 1-minute AE format (`https://wdc.kugi.kyoto-u.ac.jp/aeasy/format/aeformat.html`). The data plot and download service (`https://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html`) does not currently support the real-time AE index but will in the future.

We have also published real-time digital values for past periods prior to December 2024. Currently, data covering the period from January 2021 to November 2024 are available.

Overview of Data processing

The calculation method for the real-time AE index (including the published digital values) is basically the same as before (see `https://wdc.kugi.kyoto-u.ac.jp/aedir/ae2/onAEindex.html` and `https://wdc.kugi.kyoto-u.ac.jp/wdc/pdf/AEDst_version_def_v3.pdf`). Before publishing the digital values, we perform a quick quality check on the original geomagnetic data used to calculate the AE index, based on visual inspection and suggestions from an automatic calibration algorithm. We then correct baselines and remove noise where necessary. This check and correction process is quicker than for the provisional index, and if a correction is required, suggestions from the automatic calibration algorithm are generally adopted directly. Please note that data publication may be put on hold if the data prove difficult to correct.

Once the digital values have been published, the data and their plots on the website will not be updated, even if the original geomagnetic data are subsequently updated. If you wish to use a higher-quality AE index, please wait for the provisional index to be published.

We currently do not use the average values of the international five quietest days (IQDs) as the baseline of geomagnetic data used for the real-time AE calculation. Since the international five quietest days are published after the end of each month, determining the baseline in near real-time using this method is problematic. Therefore, the baseline is determined on an as-needed basis, rather than monthly, based solely on the aforementioned data correction process, regardless of whether baseline changes are due to natural or artificial variations.

Baseline Determination Algorithm

The algorithm estimates the baseline for a target day using H-component (or relevant component) geomagnetic data from a 5-day window centered on that day. The core processing involves the following steps:

1. **Data Segmentation and Statistics:** The data within the 5-day window are divided into 3-hour intervals. For each interval containing sufficient data (>75%), the median value and the interquartile range (IQR) are calculated. The median serves as a robust estimator of the central magnetic field value, minimizing sensitivity to transient disturbances like substorm bays or spike noise. The IQR, calculated after removing a linear trend within the interval, quantifies the data variability and serves as an indicator of geomagnetic activity or noise level.

2. **Weighted Polynomial Fitting:** A polynomial function representing the baseline trend is fitted to the sequence of 3-hourly median values (y_i) across the 5-day window using a weighted least-squares (chi-squared minimization) method. The weight assigned to each median value is determined by the inverse square of its corresponding IQR (σ_i), effectively reducing the influence of data from disturbed intervals (high IQR). The minimized objective function is:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{y_i - y(x_i)}{\sigma_i} \right)^2$$

   where $y(x_i)$ is the polynomial value at time x_i.

3. **Adaptive Polynomial Order:** The order of the fitted polynomial is adapted based on the contiguous availability of data within the 5-day window: 3rd order for 5 days, 2nd order for 3-4 days, and 1st order (linear) for 1-2 days.

4. **Baseline Output and Formatting:** The fitted polynomial, derived from the 5-day window analysis, is evaluated specifically for the central target day (the day for which the baseline is being estimated) to constitute the preliminary continuous baseline value for that day. To ensure compatibility with manual correction procedures, this continuous baseline undergoes a two-step quantization process. First, the daily baseline profile is scaled or adjusted such that its minimum and maximum values for the day align with multiples of 20 nT. Second, the resulting (scaled/adjusted) baseline values for each minute are rounded to the nearest 20 nT multiple. If data availability within the window is severely limited, the final baseline value from the preceding day is propagated forward.

5. **Monthly Constant Offset Determination and Application:** As a final refinement, a constant offset for the entire month is determined, with 1 nT resolution. For near real-time processing, this calculation occurs at the beginning of the following month using data from the preceding month. However, for past data prior to December 2024, the offset for a given month is calculated using data from that same month. The calculation utilizes the data already adjusted by the daily varying baseline component (from step 4). Specifically, 3-hourly median values are computed from this adjusted data for the entire relevant month (preceding or current, depending on processing mode). A weighted average of these median values is then calculated, using the corresponding IQR of the 3-hourly segments as weights (inversely squared). This weighted average represents the constant monthly offset (effectively a 0th-order weighted polynomial fit). This offset, typically less than 10 nT and does not exceed 20 nT, is then added to the daily varying baseline component for all data points within the target month to yield the final baseline.

This automated approach facilitates daily baseline determination suitable for near real-time AE index production. By utilizing the median and IQR-based weighting, the method demonstrates robustness against transient geomagnetic events and provides a baseline that dynamically adapts to slow variations. Comparisons indicate that the automatically determined baseline generally aligns within ±50 nT of baselines derived from manual inspection or IQD methods for the vast majority (>99%) of intervals, though larger deviations may occur during periods of sustained, intense geomagnetic activity or adjacent to significant data gaps. An example illustrating the performance of the automatic baseline correction compared to visual correction is shown in Figure 1.

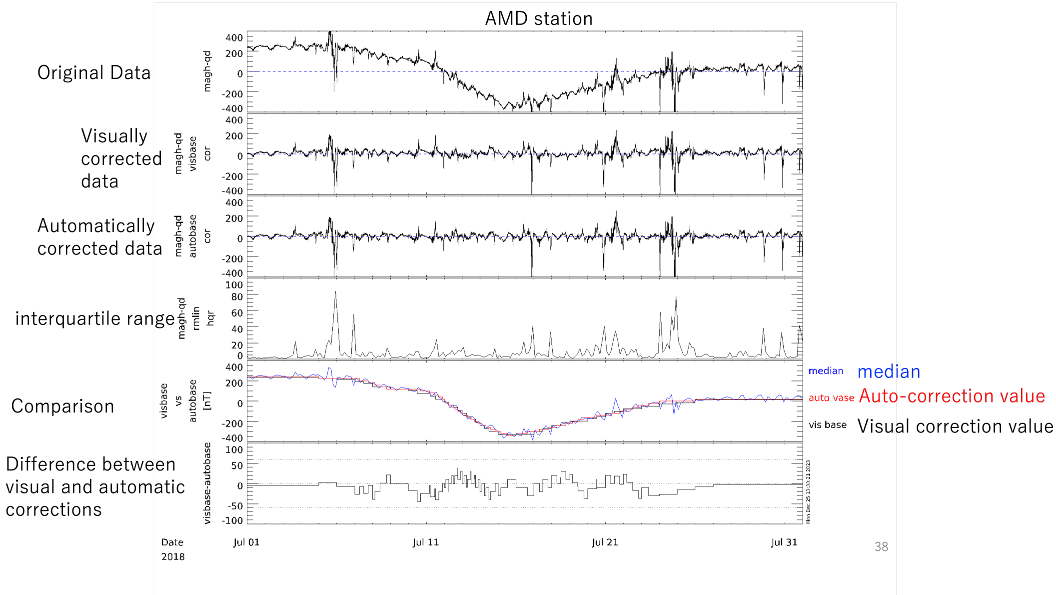

Figure 1: Example of automatic baseline correction for AMD station data (July 2018). Panels show (top to bottom): original data (relative to quiet day baseline), visually corrected data, automatically corrected data, interquartile range (IQR), comparison of estimated baselines (blue: 3-hourly median, red: automatic correction value, black: visual correction value), and the difference between visual and automatic baseline correction values. All vertical axes are in units of nT.

Spike Noise Determination Algorithm

The spike detection algorithm processes the 1-minute geomagnetic H-component data for each AE station independently. The core steps are as follows:

1. **Calculate Point Deviation ("Sharpness"):** A measure quantifying the sharpness or deviation of each data point (B_t) relative to its adjacent points (B_{t-1}, B_{t+1}), denoted here as D_t, is calculated. This captures the magnitude of rapid, isolated changes characteristic of spikes, potentially related to the second difference like $D_t = |B_t - (B_{t-1} + B_{t+1})/2|$.

2. **Estimate Local Variability:** The local variability of the signal is estimated using IQR, denoted as IQR_t, calculated from the magnetic field data (B_t) within a moving time window (e.g., 30 minutes). This variability calculation is performed after removing a linear trend from the data within the window, quantifying the typical magnitude of fluctuations independent of slow baseline drifts.

3. **Stabilize Variability Measure:** To prevent division by very small numbers during quiet intervals and ensure numerical stability, a small constant, denoted as C (e.g., $C = 3$ nT), is added to the calculated local variability measure (IQR_t).

4. **Compute Normalized Spike Index:** A normalized index (S_t) quantifying the "spike-likeness" of each point is computed by dividing the point's sharpness measure (D_t) by the stabilized local variability measure ($IQR_t + C$):
$$S_t = \frac{D_t}{IQR_t + C}$$
This index represents the magnitude of a point's deviation relative to the expected local fluctuation level.

5. **Thresholding and Flagging:** Data points are flagged as potential spike noise if their calculated normalized spike index (S_t) exceeds a predefined threshold. Analysis suggests that values of S_t greater than approximately 25 often correspond to visually identified spikes, distinguishing them from natural variations.

Contact

Contact on the digital value of the real-time AE indices: Shun Imajo (wdc-service@kugi.kyoto-u.ac.jp)