Quality Control Procedures At The World Data Centre for Geomagnetism (Edinburgh)

Ewan Dawson (ewan@bgs.ac.uk), Sarah Reay, Susan Macmillan, Simon Flower, Tom Shanahan
British Geological Survey, Edinburgh, UK

Abstract

Geomagnetism data holdings at the World Data Centre for Geomagnetism (Edinburgh) include observatory minute, hourly and annual mean values and global magnetic survey and repeat station data. Until 2007 the observatory minute and hourly mean value holdings were maintained by the Danish Meteorological Institute. We describe the quality control procedures developed at the British Geological Survey for application to new and legacy data.

A number of issues raised for the legacy observatory hourly mean values result from formatting and typographical mistakes. The clearly obvious mistakes are being corrected in the WDC data that are made available at www.wdc.bgs.ac.uk. Whilst these quality control procedures have been mostly developed with the application of global modelling in mind, they will have benefits for all users of the data. Some examples of corrections are given.

Data quality control policy

The core principle of the WDC data quality control policy is that the WDC should not make any changes to the data it receives, beyond the correction of obvious formatting or typographical errors. As such, the WDC does not perform extensive quality control processing such as spike identification and removal or baseline shifting. If, in the course of our quality analysis, any such suspected errors are found, the WDC will notify the contributor of the data. The WDC may work with the contributor to improve the quality of the data, but it is the responsibility of the contributing institute to make corrections and re-publish the data.

Should our analysis identify any obvious typographical or formatting errors, the WDC will make the necessary corrections, and make the edited data available for download via the WDC website.

We do not specify a minimum data quality acceptance criteria, beyond that of requiring the data to be produced by an observatory-standard, IAGA-recognised installation.

Software used for data quality analysis

Data is analysed using the INTERMAGNET CD-ROM Viewer v1.3, (IMCDVIEW) which is produced by the British Geological Survey and distributed by INTERMAGNET. This software provides a convenient way to inspect minute, hourly and daily mean data (see Figure 1).
We also use software to plot hourly mean values over larger timescales (see Figure 3).
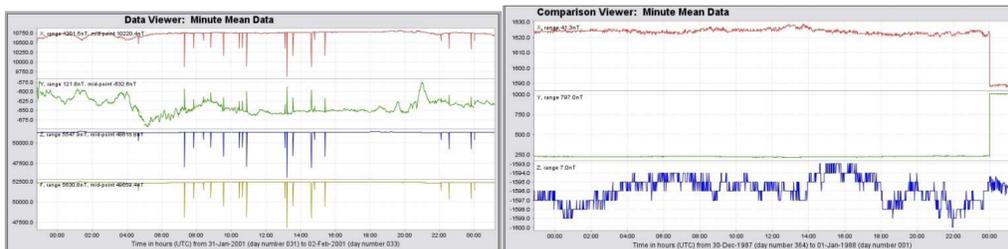

Figure 1: Screenshots of INTERMAGNET CD-ROM viewer, showing errors in minute mean data (left), and a step identified by comparison with a nearby observatory (right).

Quality control procedure

On receipt of minute means values dataset:

We first transform the data from the original format (usually WDC or IAGA-2002) into INTERMAGNET CD binary format. This allows us to examine the data using IMCDVIEW. We then plot the data on yearly, monthly and daily timescales, examining the plots to identify data with potential quality issues. If possible, we then compare the data with that of a nearby observatory for the same period. IMCDVIEW has a useful function to plot the difference magnetogram for nearby observatories. This is especially useful in identifying any spikes, intra-year jumps and drifts in the data.
If we do not hold hourly or annual means from the observatory for the same period, we compute these data from the minute data.

On receipt of hourly mean values dataset:

We transform any non-WDC format data into INTERMAGNET CD binary format and analyse as described above. For data submitted in WDC format, we plot the monthly means at that observatory for recent years (including the newly submitted data) to detect formatting problems, e.g. incorrect tabular base values and missing signs, and for intra- and inter-year jumps. These kinds of problems are especially common in WDC-format hourly mean data, due to the relatively complicated nature of the data format.
Again, annual means are computed from the hourly mean data if not already available.

On receipt of annual mean values dataset:

We plot annual means and annual rate of change, with known jumps applied, to detect unspecified inter-year jumps. The internal consistency of the data is also checked, e.g I = arctan Z/H.

If there are any data quality issues thrown up by these procedures, we query them with the data provider. Otherwise the data are made available at www.wdc.bgs.ac.uk.

Quality control of WDC hourly mean files

The majority of hourly mean data submitted by non-INTERMAGNET observatories are in WDC hourly mean value format. An example of data in WDC hourly mean value format is given in Figure 2. Each hourly mean value is decomposed into a tabular base value (shared by all mean values for a given day), and a tabular value.

Tabular bases   D, H & Z mean values for hour 1, day 1, month 1, 2008 are 345 deg 47.0 min, 12575 nT and 50161 nT

```
LRV0801D01     20 345 470 436 429 418 425 447 459 447 437 434 432 423 413 411 404 412 415 418 423 427 433 440 462 441 432
LRV0801D02     20 345 440 437 436 435 430 435 428 433 436 437 436 432 418 406 403 406 415 412 415 423 429 434 439 443 427

LRV0801H01     20 120 575 571 584 582 589 592 595 595 595 593 591 589 589 589 589 591 593 592 593 592 590 590 589
LRV0801H02     20 120 592 586 585 588 590 592 594 596 595 596 593 590 588 587 588 591 594 597 595 592 594 591 591 588 591

LRV0801Z01     20 500 161 144 148 166 179 180 177 176 179 180 181 182 183 183 185 186 186 185 185 184 185 188 189 182 178
LRV0801Z02     20 500 182 181 182 184 183 182 183 182 181 182 182 182 183 183 184 185 185 185 188 191 189 189 186 184 184
```

Figure 2: WDC hourly mean value file format, with tabular base value highlighted.

We are currently in the process of checking our entire dataset of legacy annual mean values for typographical and formatting errors. Files of observatory hourly mean values for 1968-2008 held in WDC format have already been checked. Some examples of the kind of errors found are:
- Sign of tabular base incorrect.
- Tabular base of two components switched.
- Component incorrectly identified in column 8.
- Tabular base value incorrect.
- Sign of tabular values incorrect.
- Incorrect missing value flag used (e.g. 999 instead of 9999).

These kinds of errors are all relatively easy to spot and correct.

Legacy datasets corrected to date:

- Annamalainagar (ANN) 1967-1979
- Arti (ARS)   1977, 1995-1996
- Bangui (BNG) 2004
- Eskdalemuir (ESK) 1983-1989
- Fuquene (FUQ)   1984-1989, 1991-2000, 2002-2006
- Hartland (HAD) 1983-1989
- Hurbanovo (HRB) 1987, 1992, 1994
- Irkutsk (IRT) 1991-1994
- Istanbul Kandilli (ISK) 1989-1994
- Bereznyaki (KGD) 1975
- Lerwick (LER) 1983-1989
- Magadan (MGD) 1987
- Valencia (VAL) 1993
- Neumayer Station (VNA) 2008
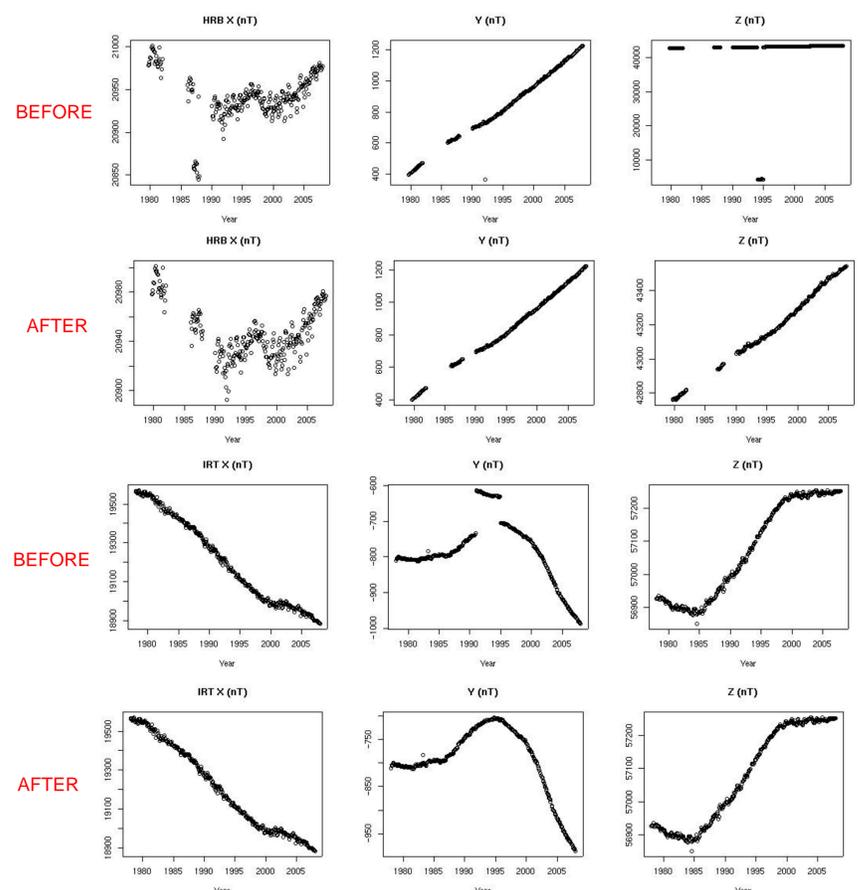- Yazhno Sakhalinsk (YSS) 1981-1985

Figure 3: Plot of monthly mean values before and after correction.