OBSERVATORY DATA QUALITY CONTROL – THE INSTRUMENT TO ENSURE VALUABLE RESEARCH

H.-J. Linthe [(1)], J. Reda [(2)], A. Isac [(3)], J. Matzka [(4)], C. Turbitt [(5)]

[(1)] GFZ Potsdam - Adolf Schmidt Geomagnetic Observatory, Lindenstr. 7 14823 Niemegk Germany, linthe@gfz-potsdam.de
[(2)] Institute of Geophysics, Polish Academy of Sciences, 01-452 Warsaw, Poland, jreda@igf.edu.pl
[(3)] Geological Institute of Romania, No. 1 Caransebes Str. Sector 1 RO-012271 Bucharest, Romania, margoisac@yahoo.com
[(4)] Technical University of Denmark - National Space Institute, Elektrovej 327, 2800 Kgs. Lyngby, Denmark, jrgm@space.dtu.dk
[(5)] British Geological Survey, West Mains Road Edinburgh EH93LA, UK, cwtu@bgs.ac.uk

SUMMARY

Observatory data are the foundation for international scientific research. Valuable results can be achieved only if the data are precise and faultless. High quality instruments and a high level of ability and motivation of the observatory personnel are necessary, but a rigorous process of checking is as important for data quality control. Observatory data are useful for science only if the quality can be assured through peer review prior to publication.

INTERMAGNET encourages participating observatories (IMOs) to check their definitive data before they are submitted. Furthermore, definitive data are carefully double-checked by volunteers and by the Definitive Data Subcommittee before they are published to the scientific community. This procedure is labour-intensive, but is essential to maintain a consistently high level of data quality.

Different methods of data checks are described and their efficiency is discussed with consideration to the different instrument base of an observatory. Reason will be given as to why each data check is necessary and the tools available for effective data checks are described.

1. INTRODUCTION

Since Gauss' time in the early 19th century, geomagnetic observatories have monitored the Earth's magnetic field following the same principle. As a result of limitations on instrumentation, all observatories make separate absolute measurements and variometer recordings i.e. variometer data need to be calibrated by means of absolute measurements. Final observatory data are a combination of absolute measurements and variometer recordings. Many aspects of data quality ensuring can be found in Reda et al. (2011).

2. ABSOLUTE MEASUREMENTS

Variometer recordings are periodically calibrated to the absolute level by means of manually performed absolute measurements, which estimate the variometer baselines i.e. the non-constant difference between a variometer output and the absolute field vector. Since baseline values are usually isolated spot values, the deviation of a variometer's baseline between absolute measurements is modelled by interpolation (typically polynomials or splines). Today's standard instrument set consists of a fluxgate-theodolite, used to determine the declination and the inclination of the field vector, and a scalar magnetometer for the measurement of total intensity.

The quality of the measurement results depends on the quality of the instruments and on the ability of the observer making the measurements. For all measurements, the measurement environment is critical and instruments should be operated in a controlled, magnetically clean facility. Since fluxgate-theodolite measurements are currently a manual procedure at most observatories, observer training and a well-defined practice are key to good quality absolute measurements. The biennial IAGA Workshops on Geomagnetic

Observatory Instruments, Data Acquisition and Processing, provide a means of standardising instruments and observer practice internationally through measurement comparisons at a single observatory.

3. VARIOMETER RECORDINGS

Because the Earth‘s magnetic field is a vector, a three-component variometer is required to record its time-varying intensity and direction. A number of differing variometer technologies have developed over the last 180 years; many observatories now operate triaxial fluxgate magnetometers while others operate triaxial photoelectric feedback magnetometers based on a suspended magnet. Other observatories are using vector magnetometers based on scalar magnetometers with compensation coils (for instance, the dIdD variometer).

Baseline stability and effective sampling of any baseline drift through good quality absolute measurements are important quality criteria for an observatory. Such baseline drift may be caused by temperature changes in the variometer room or by tilting of the measuring pillar supporting the instrument. Such baseline drifts can be minimized by a temperature stabilized variometer room (thermal insulation, thermoelectric temperature control) and sensor suspension.

Modern variometers record the variations of the Earth’s magnetic field in digital form and the most straight-forward quality check is a regular inspection of the plotted recordings. A simple daily magnetogram plot of all components can readily show serious problems, such as large scale instrument noise, external interference and data outages. However, a plot of ΔF (difference of F calculated from the baseline corrected vector variometer minus F continuously recorded by a scalar magnetometer) is much a more sensitive detector of instrument problems such as baseline and scale value errors, temperature drift and anthropogenic noise. This ΔF method of problem identification is only possible where an observatory is making continuous F recording by means of a separate scalar magnetometer. The parameters detectable using the ΔF method depend on the type of variometer being used e.g. because coil-compensated scalar magnetometers use a similar F magnetometer, parameters such as baseline and scale errors or temperature drift cannot be checked by additional scalar recording.

If the observatory operates more than one variometer set, the recordings should be compared with each other in order to provide additional quality information. Such comparisons can identify component-specific problems such as scale value errors or, where the variometers are not co-located, room temperature instabilities, data logger time synchronization problems and artificial noise.

4. DATA MANAGEMENT

In addition to good measurement practice, a significant amount of data processing is required to produce a definitive time-series from a set of variometer records and absolute measurements. To maintain the quality level required by scientific research, observatory data require accurate and reliable data management procedures through, inspection, correction, baseline derivation, to preparing for publication. Because the global observatory data set spans such a large period of time, practice and procedure need to be rigorous and consistent in order to ensure that this data set is homogeneous, hence observatory staff are required to be well trained and experienced. This becomes more and more important in a time of an increasing number of unmanned observatories, which are operated by a remote institute. It is currently common place that a single institute will operate a network of geographically disparate observatories with the responsibility for absolute measurements contracted to a third party. Operating such an observatory model brings new challenges in maintainging the measurement environment, training observers and meeting data quality standards.

In the frame of the International Geophysical Year, the system of World Data Centres (WDCs) was established during the 1950s. This improved the use of observatory data. During the 1970s and 1980s, the number of magnetic observatories making digital rather than analog recordings increased rapidly and the WDCs in turn adapted their data management procedures to this trend.

At the end of the 1980s a new data collection and distribution system was established initiated by the British Geological Survey (BGS), Canadian Geological Survey (CGS), the Institute of Physics of the Earth Paris (IPGP) and the United States Geological Survey (USGS). Over the subsequent years, INTERMAGNET became a global network of observatories, taking advantage of near real-time data dissemination and the quality control offered by digital data.

5. INTERMAGNET

Although INTERMAGNET was established using a limited number of magnetic observatories, it was quickly extended to include all absolute magnetic observatories capable of meeting set data quality and distribution standards:

- Near real-time distribution of Reported Data within less than 72 hours. Reported Data are unchecked variometer data, not baseline corrected or only approximately corrected.

- Supply of Definitive Data, which meet a specified data quality standard and which are comprehensively checked for their quality by INTERMAGNET.

With time, an increasing number of international observatories have attained status of an INTERMAGNET observatory (IMO). As of 1 March 2013, 133 observatories are part of the INTERMAGNET network. Besides Reported and Definitive Data, a new data product has recently been established to meet the demand by the magnetic field modelling community for good quality data, delivered more rapidly than definitive data: Quasi-Definitive Data are close to the expected Definitive Data, de-spiked and baseline adjusted, but to be delivered within three months rather than annually**.**

In the beginning Reported Data were distributed by satellite transmission, and Definitive Data were provided on CD-ROMs. Nowadays the web is the base of the data submission and distribution. Reported data and Quasi-Definitive Data are transmitted by e-mail or by a web based data delivery system to one of the 5 Geomagnetic Information Nodes (GINs) and then to the INTERMAGNET web site. Definitive Data are now available at the INTERMAGNET web page for download and on DVDs. Reported and Quasi-Definitive Data can be also downloaded from the INTERMAGNET web page.

INTERMAGNET requires that the IMOs to meet a specified data quality standard and provides support to observatories in the form of software for data processing; quality and data format checking; training for observers on request; and assists IMOs in the processing data from the raw recorded stage to the final published product.

One of the principle tools made available by INTERMAGNET for data quality inspection has been created by the BGS – the so called CD (or DVD) Data Viewer. This is not only a simple data viewer, but it is a comprehensive instrument to detect data problems. In addition to plotting 1-minute data, the application can be used to view hourly, daily, annual mean values and variometer baselines, thus providing analysis of the long-term performance of the data set. It is also possible to plot the first differences of 1-minute data and comparisons (differences) of 1-minute data against neighbouring IMOs. The plotting utility has a number of features that are tailored to monitoring the continuous, long-term nature of observatory recordings e.g. the time window can be extended to plot part of the previous day, month and year and the following one, respectively.

The functionality of the INTERMAGNET CD Data Viewer is particularly useful for an IMO preparing Definitive Data for publication. For example, being able to compare one observatory time-series against another nearby observatory can detect artificial noise generated at a distance to one of the observatories. This type of noise (where both scalar and vector instruments at an observatory see the same level of noise) will be invisible in the ΔF plot of a single observatory. In addition, the ΔF plot will not detect any perturbation affecting only the declination component as the East component generates only a very weak influence on ΔF. As stated previously, observatories operating exclusively variometers based on scalar magnetometers with compensation coils (e.g. dIdD variometers) can often only detect problems in the orthogonal components through comparison with the data of neighbouring observatories.

Plots of first differences plot using the CD Data Viewer can be especially useful in assessing the artificial noise level in a given observatory, particularly for transient interference, such as steps or spikes, at levels above the natural signal.

Baseline jumps or drifts can be found also by means of a comparison plot against data from neighbouring observatories. Baseline variations are typically low amplitude and long period so often, baseline values are assigned on a daily basis. For magnetograms plotted in time windows of 24 hours, baseline jumps between consecutive days can go undetected but if the time window is extended by some percent into the previous and the following day, baseline jumps can become evident.

It is worth noting here that special care should be taken to avoid misinterpretation. In some cases natural variations can be wrongly recognized to be artificial noise. In cases of a suspected artificial disturbance, it is necessary to look carefully at the magnetograms of neighbouring observatories, their first differences plots and at the difference plots. Short periodic natural variations such as sudden storm commencements (SSC) can manifest as irregularities in the ΔF plot due to differing sampling rates, filtering and aliasing. 1-minute vector data are routinely processed from higher frequency data samples using a Gaussian filter while the scalar data can be one minute samples following the present INTERMAGNET standards, given in St. Louis (2011).

One of INTERMAGNET's current projects is planning for the distribution of 1-second data. The first steps for data quality validation have already been undertaken in that the data quality parameters, such as the accuracy of the time stamp and noise characteristics, have been specified. Developing processes to ensure that observatory data meets these specifications and the processes involved in checking these data are the next challenge for the INTERMAGNET team.

6. MAGNETIC CLEANLINESS

It is obvious to state that a magnetically clean environment is most important for the successful operation of a magnetic observatory, but this is frequently difficult to achieve in practice. The distance from the measurement point to ferromagnetic materials and DC currents needs to be sufficient to avoid their influence. Big cities produce a high level of industrial and urban noise due to moving vehicles and electric railways. A number of long-term operated observatories have had to be moved due to expanding cities, increasing industrial activities and traffic. External noise is extremely disturbing for the successful operation of a magnetic observatory and this is normally out of the control of the observatory, however it is critical that an observatory produces no artificial noise by its own activities. Where laboratories or workshops are operated at the site of an observatory, personnel must continuously take care to ensure that variometer recordings and absolute measurements are undisturbed. The number of observatories, having such activities is small and probably further decreasing, as more and more observatories become unmanned, although it s worth noting that a disadvantage of an unmanned observatory is the loss of close supervision on activities which may be a source of artificial noise in the immediate vicinity of the observatory

Randomly occurring artificial noise can be recognized by means of the ΔF check from the vector and scalar recordings of the observatory of concern or by means of difference plots to neighbouring observatories as described in the sections 3 and 5 but noise level can be of a level that it can be seen in a plot of the first differences of a quiet day. Where there is no neighbouring observatory, a useful comparison can be that of the first differences of the same day of an observatory of nearly the same geomagnetic latitude.

Modern infrastructure, such as DC powered railway lines or electrical DC power distribution systems, often produce increasing levels of perturbation on particular magnetic observatories as demand on that infrastructure increases with time. It is, in most cases, impossible to mitigate against this noise. The economic and public significance of the operation of railway and electrical power distribution are of higher importance than the noise-free working of a magnetic observatory. In the past, many magnetic observatories have been moved to places promising (hopefully) sufficiently quiet conditions for the further long-term operation of the observatory. Such a movement is extremely expensive, time consuming and in order to achieve a significant distance from the noise source, the time series of the observatory can, in general, no longer be considered continuous. Some observatories are currently faced with this problem, or are already in planning to move, as the only open to an observatory not able to continue operating at a new location is closure, , which has also unfortunately happened in some cases.

To correct artificially disturbed observatory data series by means of mathematical modelling of the anthropogenic noise and subtracting the modelled signal from the recordings is non-trivial and is frequently not possible. Fox Maule et al. (2009a) tried to establish a useful method for the “cleaning” of the Brorfelde (BFE) time series from perturbations produced by two DC electrical power lines in the Baltic Sea. Neska et al. (2011) describe comprehensively the analysis of the effect of DC powered railway lines on Belsk (BEL) and Lviv (LVV) observatories and suggest an algorithm to correct the contaminated data. But the authors clearly state: “By the way, every trial to correct data can produce unwished secondary effects in the result. For this reason, only original data are accepted by the World Data Centers that collect the observatory data. So our reconstructed data are by no means supposed to substitute the original data.”

7. CONCLUSIONS

High quality observatory data are highly appreciated by the international geomagnetism research community. A number of observatories of exceptional position and particularly long-term time series are of extremely significant interest. Observatory data quality is not explicit– it requires experience, thoroughness and is time consuming. The production of high quality geomagnetic observatory, long-term time series is a challenge that is often not highly enough recognised by data users. Faced by staff reduction, ensuring data quality becomes more of a challenge to institutes operating observatories. Often the operation of observatories is not sufficiently acknowledged by funding organisations and agencies. The result of the painstaking work applied to observatory operation is not directly visible in a time-series alone. The routine work of observers does naturally not lend itself to the publication of scientific articles, however numbers of publications are often the criteria on the results of scientific institutes. Nevertheless, many high level scientific publications are dependent on the high quality absolute magnetic observatory data series.

8. REFERENCES

Fox Maule, C., P. Thejll., A. Neska, J. Matzka, L. W. Pedersen. and Nilsson, A. (2009a). “Analyzing and correcting for contaminating magnetic fields at the Brorfelde geomagnetic observatory due to high voltage DC power lines”. Earth Planets Space **61**, 1233–1241.

Neska, A., J. Reda, M. Neska and Y. Sumaruk (2012). “On the influence of DC railway noise on variation data from Belsk and Lviv geomagnetic observatories”. ACTA GEOPHYSICA (2012), DOI: 10.2478/s11600-012-0058-0

Reda, J., D. Fouassier, A. Isac, H.-J. Linthe, J. Matzka and C. W. Turbitt (2011). "Improvements in geomagnetic observatory data quality". Geomagnetic Observations and Models, IAGA Special Sopron Book Series 5, 127-148.

St. Louis, B. (2011): "INTERMAGNET Technical Reference Manual Version 4.5". http://www.intermagnet.org/TechnicalSoft_e.php