PUBLS. INST. GEOPHYS. POL. ACAD. SC., C-99 (398), 2007

Worldwide Observatory Hourly Means from 1995 to 2003: Investigation of Their Quality

Monika KORTE[1], Mioara MANDEA[1] and Nils OLSEN[2]

[1] GeoForschungsZentrum Potsdam
Telegrafenberg, 14473 Potsdam, Germany
e-mails: monika@gfz-potsdam.de; mioara@gfz-potsdam.de

[2] Danish National Space Center
Juliane Maries Vej 30, 2100 Copenhagen, Denmark
e-mail: nio@spacecenter.dk

Abstract

We examined the hourly mean values from worldwide geomagnetic observatories as available from the World Data Center C1, Copenhagen, for erroneous or problematic data. A number of outliers and jumps were found, as well as instabilities in baseline control. It is impossible to correct all these data without checking the original values at the individual observatories again. We provide a table of identified problems to alert both users and the data owners about them and encourage each observatory to correct their data where possible.

1. Introduction

Geomagnetic observatory data are used for a multitude of different purposes, from main field and secular variation studies to investigations of external magnetic field variations and mantle conductivity. With digital data recording the availability of high resolution geomagnetic time series has increased (Fig. 1). These data are an important complementation of the recent high resolution magnetic data obtained by satellites like Ørsted and CHAMP. Hourly mean values (HMVs) and one minute values are common data products from geomagnetic observatories (Jankowski and Sucksdorff 1996, St.-Louis 2004). The data are delivered to World Data Centers (WDC), where they are stored and freely available for scientific purposes. However, erroneous data can find their way into the databases. Sources of error are plentiful: Errors simply might occur in data transfer or importation to the database. Measurement errors happen and uncorrected values might be delivered to a WDC by mistake. Moreover, raw observatory data consist of variation measurements and baseline determinations, which have to be combined to the final absolute time series. Producing good final data requires both skill in the absolute measurements to determine the baseline and care in processing the data. Even slight mistakes in either of these steps can lead to poor baseline control, which can result in artificial variations in the time series.

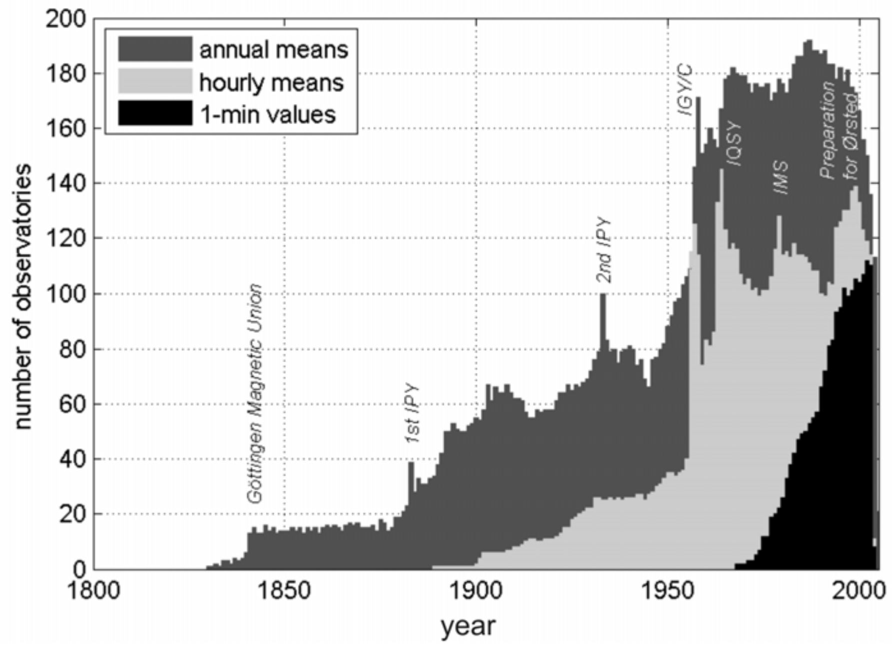


Fig. 1. Number of geomagnetic observatories providing annual means, hourly means and 1-min values, as a function of time. International campaigns, like the *Göttingen Magnetic Union*, the 1st and 2nd *International Polar Year (IPY)*, the *International Geophysical Year (IGY)*, the *International Quiet Solar Year (IQSY)*, the *International Magnetospheric Study (IMS)* and the preparation of the Ørsted satellite mission have stimulated observatory data processing and encouraged the opening of new observatories.

HMVs are well suited to detect most problems in high resolution observatory data and are widely used in scientific studies (e.g. Balasis and Egbert 2006, Le Mouel *et al*. 2005, Sabaka *et al*. 2004, Macmillan and Thomson 2003, to name just a few recent publications). We started a project to check all the available HMVs for outliers, jumps (i.e. uncorrected sudden baseline level changes), baseline drifts and other obvious problems. With approximately 150 observatories worldwide, this is a huge task. We first concentrated on the time from 1995 to 2003. Many observatories had changed to digital recording by 1995, and the number of available data (reported to the WDC) increased significantly then, during preparation for the Ørsted satellite mission. 2003 was the last epoch for which a reasonable number of global data was available at the time when the work was done. Here, we first describe the dataset and our main method of quality check. We mention further methods which we tested and which also can be useful to detect problematic data. We summarise the results of our work and show some examples of detected problems. The full list together with figures is available online at http://www.gfz-potsdam.de/pb2/pb23/COM/comparison_obs.html.

For several reasons it is beyond the scope of our work to actually correct the data. First, we are not authorised to manipulate other observatories' data. Second, in many cases a correction is not straightforward. Except for some cases where individual outliers could be removed or corrected by interpolation or where the magnitude of a clear jump might be estimated, it will be necessary to go back to original data sources to figure out the source of the problems. Doing this for all observatories is beyond our capacity. Third, if baseline problems result from insufficient or inaccurate absolute measurements, it will unfortunately very likely be impossible to correct for them. Our basic intention is to alert data producers and users to problematic data. Given the importance of high quality geomagnetic observatory time series we urge all the original data holders to make efforts to correct erroneous data and update the WDC archives.

2. Checking Data Quality by Comparison

The data we used are the annual HMV files that were available at WDC C1 Copenhagen (http://dmiweb.dmi.dk/fsweb/projects/wdcc1/) in March 2005. The data were checked by intercomparison, i.e. the annual data series from each observatory were visually compared to those of at least two neighbouring observatories. We implemented a method to plot the north (X), east (Y) and vertical (Z) components of two observatories together with the differences. Examples are shown in Fig. 2 to 4. All these figures show real observatory data, but the stations have been made anonymous because similar features as shown here can be seen in data series from several other observatories, too.

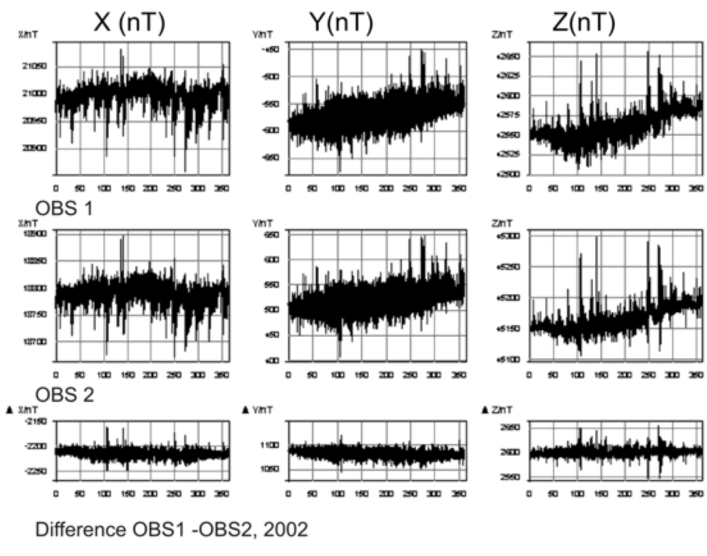


Fig. 2. Comparison of hourly mean values from two observatories. Example for a distance of a few hundred km and no detected data problems.

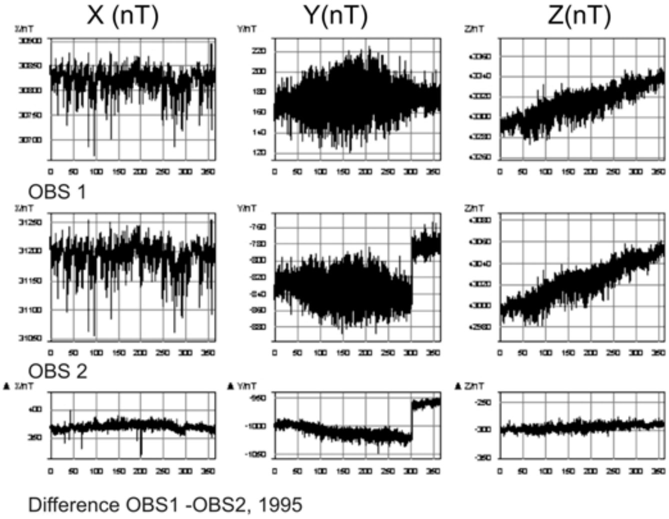


Fig. 3. Comparison of hourly mean values from two observatories where a baseline jump is clearly seen in Y and less clearly in X at about day 300 of the year.

In the differences, we expect to see constant offsets due to core and lithospheric field morphology, slow deviations due to differences in local secular variation, and high-frequency scatter due to differences in external and induced field contributions. Differences in secular variation and external fields depend on the distance between stations, and at times of higher levels of geomagnetic activity the differences in external and induced variations are more pronounced. These features can be seen in Fig. 2, where two observatories about 800 km apart are compared for a year when no data problems have been detected.

Spikes above the general high-frequency variation level which occur during magnetically quiet to moderately disturbed times, sudden jumps or strong variations on time-scales of a few weeks to months in the differences, on the other hand, can indicate data errors and problems in baseline control. Fig. 3 shows an example where a baseline jump is obvious in Y at about day 300 of the year. In Y the jump even is obvious in the individual data series from the second observatory. Fig. 4 shows an example where several problems are detected. This demonstrates how the comparison to two other observatories, respectively, helps to determine in which of the data series the problem occurs. In the first comparison between OBS1 and OBS2 a baseline drift in X and Z, something that looks like a kind of technical disturbance, a probable jump in Y and some outliers are marked. In the second comparison between OBS3 and OBS2, most of the problems are seen again except for two of the outlier spikes. Consequently, all problems seen in both comparisons must result in the data series common to both, in this case OBS2, and the other two outliers must originate in the OBS1 time series (which can be subsequently confirmed by a comparison of OBS1 to another observatory). In order to determine more exactly the day of the year where the problem occurred, we zoomed in on the plots. Thus we obtained a list of identified problems.

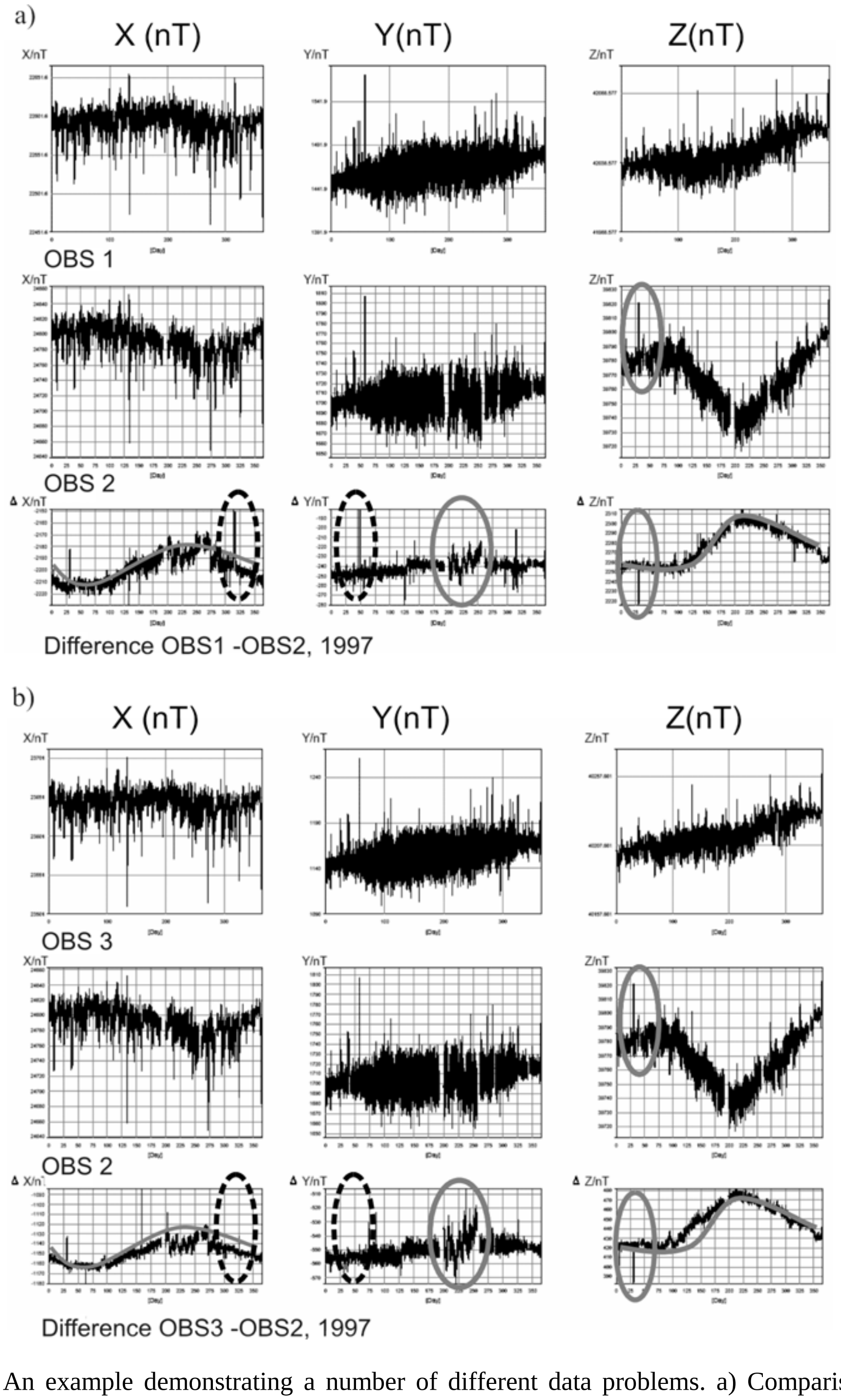
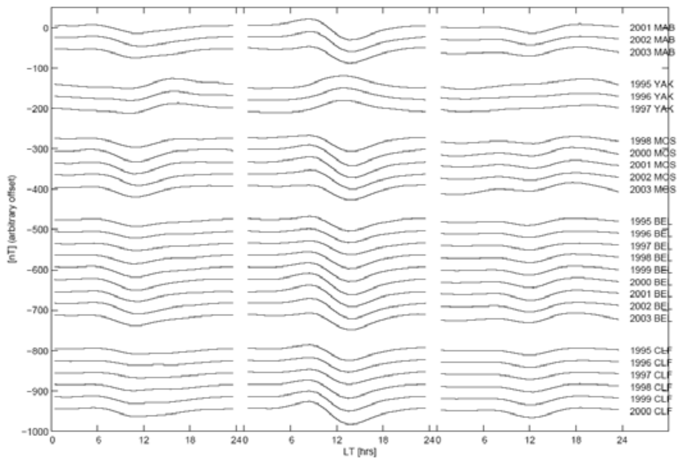


Fig. 4. An example demonstrating a number of different data problems. a) Comparison between OBS1 and OBS2, b) between OBS3 and OBS2. The problems in gray originate in the OBS2 data series, the ones marked by a dashed line originate in the OBS1 data series.

3. Further Methods

We tested two further methods of observatory data quality check, which can be performed for individual observatories without the need for direct intercomparison with others. The first is to check the mean daily variation of the components per year. It is useful to identify problems in reporting the data, particularly if data are reported with the wrong sign or in local time (LT) instead of universal time (UT). Fig. 5 shows two examples. In Fig. 5a the variations in X and Y look different at YAK than at other

a)



b)

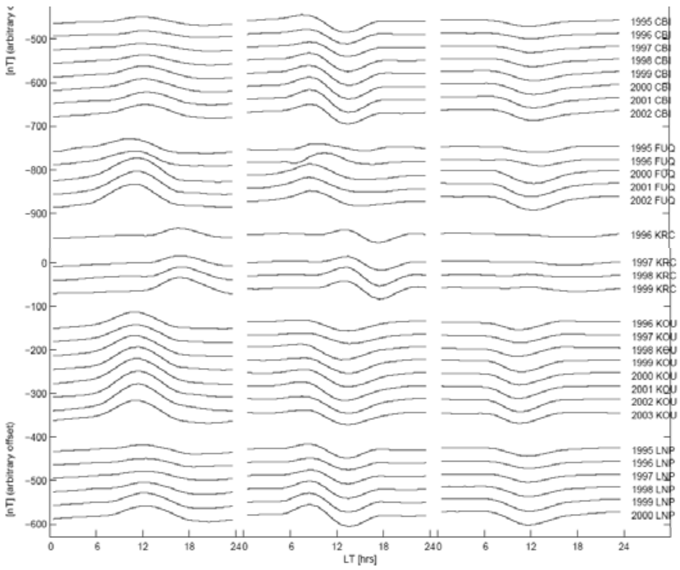


Fig. 5. Mean daily variation per year and observatory for the X, Y and Z (left to right) components. Observatories are ordered by geomagnetic latitude. Differences from the general patterns in YAK (a) and KRC (b) suggest an error in declination sign and reported time, respectively (see text).

observatories of similar geomagnetic latitude. It seems that declination recordings have been considered with the wrong sign in the determination of the X and Y component data throughout all the years. In Fig. 5b it seems that the KRC data are reported in local time instead of UT. If these data are shifted by the LT–UT difference, the mean daily variations agree much better with those of other observatories at comparable geomagnetic latitudes. Only few problems of this kind were detected. Apart from the examples shown in Fig. 5, we found that the data from observatory TTB seem to be reported in local time and in PEG this might be true for 1999.

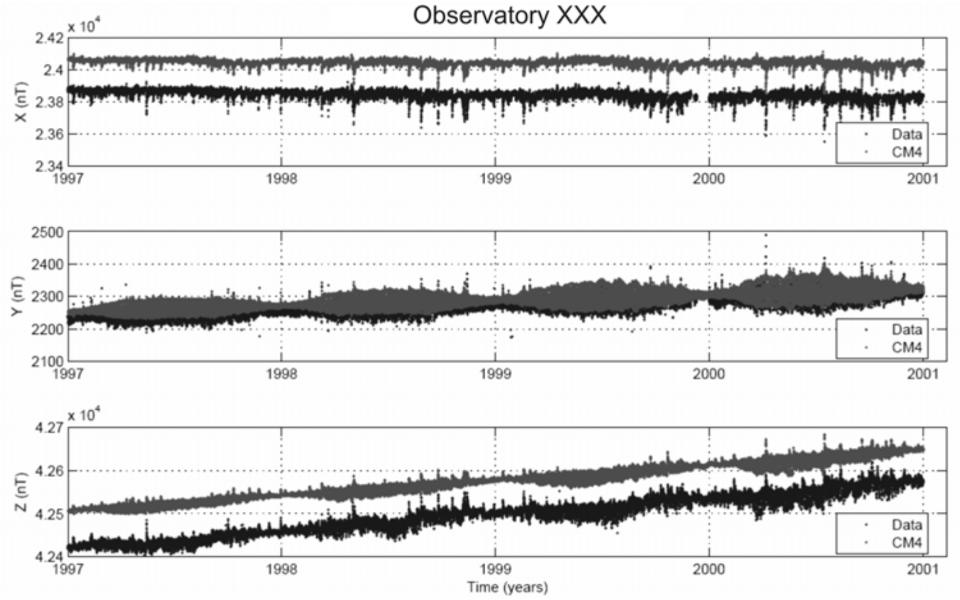
The second alternative method is to compare the observatory data to predictions of the global CM4 model (Sabaka et al. 2002, 2004). This model, covering the time span from 1960 to 2002, describes not only the geomagnetic core field and secular variation, but includes the long-wavelength lithospheric field and quiet time external fields. Modulated by the Dst index as a measure for magnetospheric activity and the F10.7 solar irradiance index, the CM4 model can also represent the variations due to external field activity to a certain degree. The geomagnetic activity is not fully described by the model, and comparisons between CM4 predictions and observatory recordings at any location look similar to comparisons between neighbouring observatories. By studying the differences jumps, outliers or baseline variations become apparent in the same way as in the comparison to another observatory. An example where a potential baseline instability is detected is shown in Fig. 6. We did not yet apply this method to all the observatory data, but we are considering to include this comparison in our future work within this project.

4. Results and Conclusions

The hourly means data series from 1995 to 2003 from geomagnetic observatories around the world, available at WDC C1 Copenhagen, have been checked for problematic data mainly by intercomparison of neighbouring stations. A number of problems have been identified, ranging from outliers in the data and jumps in the time series to continuous baseline variations. There are in fact only few observatories where we did not detect any questionable data within the 8 years that we studied. The long list of identified problems is not included in this report, but is published on a website together with plots of the intercomparisons. It can be found at http://www.gfz-potsdam.de/pb2/pb23/COM/comparison_obs.html.

It should be kept in mind, that problems are identified easily when observatories are not too far from another and lie at low to mid latitudes. At high latitudes, where geomagnetic external variations are strong, only very severe outliers or jumps can be detected with any of the methods we tested. Jumps or outliers of only a few nT are not detectable within the strong and highly inhomogeneous variations. The same is true for very remote observatories when the method of observatory intercomparison is used.

Erroneous ouliers might simply be rejected from the time series, but to correct for jumps and baseline variations it is advisable to consult the original data and check whether some real assessment of the problem can be made instead of only an estimation of the problem. Therefore it is beyond our capacity to offer corrections for the a)



b)

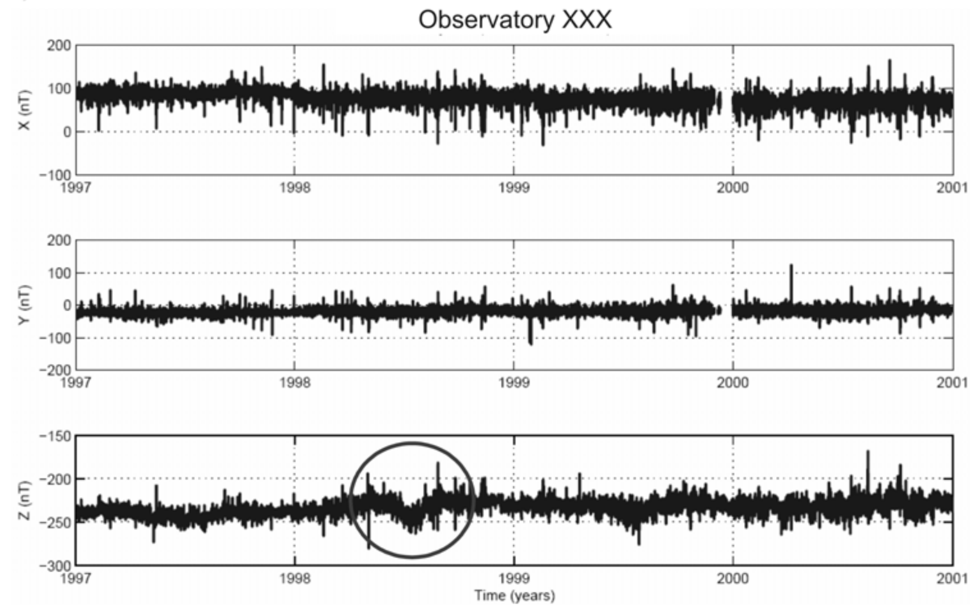


Fig. 6. a) Comparison of measured data and CM4 model predictions for the X, Y and Z component of an observatory for 4 years, b) Differences between data and model predictions. An unusual variation in the Z component in 1998 hints at a baseline problem.

identified problems. The intention of this work is to alert any users to existing problems in individual data series, but leave it to them how to handle them – whether to reject data or estimate a correction. Our main goal is to bring existing problems in the data series stored at the WDC to the attention of the producers and owners of the data, i.e. the individual observatories. We encourage them to make every effort to correct their data and update the WDC database. The long data series of high resolution observatory data still hold significant potential for research, but it is important that the data are of high quality. Meanwhile we learned that efforts are going on in France, the USA, China and India to check and improve their observatories data series.

So far we focussed on the time interval 1995 to 2003, for which many observatories have been recording the data in digital form. We will continue this project and next we will focus on the preceding time interval, from 1957 to 1994.

Acknowledgments. We thank all the geomagnetic observatories around the world that deliver their data to the World Data Centers, thus making the data easily available for all kinds of scientific studies. This work would not have been possible without preparatory data visualisation and comparison by Hans-Joachim Linthe, Jürgen Haseloff, Jutta Schulz, Katrin Tornow and Anneleen Glodek at Niemegk observatory.

References

Balasis, G., and G.D. Egbert, 2006, *Empirical orthogonal function analysis of magnetic observatory data: Further evidence for non-axissymmetric magnetospheric sources for satellite induction studies*, Geophys. Res. Let. **33**, L11311, doi:10.1029/2006GL025721.

Jankowski, J., and C. Sucksdorff, 1996, *IAGA Guide for Magnetic Measurements and Observatory Practice*, International Association of Geomagnetism and Aeronomy, 232pp.

Le Mouel, J.-L., V. Kossobokov and V. Courtillot, 2005, *On long-term variations of simple geomagnetic indices and slow changes in magnetospheric currents: The emergence of anthropogenic global warming after 1990?*, Earth Planet. Sci. Let. **232**, 237-286.

Macmillan, S., and A. Thomson, 2003, *An examination of observatory biases during the Magsat and Ørsted missions*, Phys. Earth Planet. Int. **135**, 97-105.

Sabaka, T., N. Olsen and R.A. Langel, 2002, *A comprehensive model of the quiet-time, near-Earth magnetic field: phase 3*, Geophys. J. Int. **151**, 32-68.

Sabaka, T., N. Olsen and M.E. Purucker, 2004, *Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data*, Geophys. J. Int. **159**, 521-547.

St.-Louis, B., 2004, *INTERMAGNET Technical Reference Manual*, INTERMAGNET.

Accepted February 12, 2007