Observatory data and the Swarm mission

S. Macmillan[1] and N. Olsen[2]

[1]*British Geological Survey, Murchison House, Edinburgh, EH9 3LA, U.K.*
[2]*Technical University of Denmark, Elektrovej, 2800 Kgs. Lyngby, Denmark*

The ESA Swarm mission to identify and measure very accurately the different magnetic signals that arise in the Earth's core, mantle, crust, oceans, ionosphere and magnetosphere, which together form the magnetic field around the Earth, has increased interest in magnetic data collected on the surface of the Earth at observatories. The scientific use of Swarm data and Swarm-derived products is greatly enhanced by combination with observatory data and indices. As part of the Swarm Level-2 data activities plans are in place to distribute such ground-based data along with the Swarm data as auxiliary data products. We describe here the preparation of the data set of ground observatory hourly mean values, including procedures to check and select observatory data spanning the modern magnetic survey satellite era. We discuss other possible combined uses of satellite and observatory data, in particular those that may use higher cadence 1-second and 1-minute data from observatories.
Key words: Magnetic survey satellites, geomagnetic observatory data.

1. Introduction

Magnetic observatories carry out continuous and accurate monitoring of the strength and direction of the Earth's magnetic field over many years, making measurements at least every minute. Observatory data reveal how the field is changing on a wide range of time scales from seconds to centuries, and this is important for understanding processes both inside and outside the Earth. There are approximately 160 observatories currently operating around the world. The distribution of observatories is largely determined by the location of habitable land and by the availability of local expertise, funds, data transmission infrastructure and energy supply, and as result, it is uneven and somewhat sparse in the southern hemisphere and oceanic regions.

Geomagnetism is a cross-disciplinary science, and as a result, observatories are run by a wide variety of institutes whose interests range from geology, mapping, geophysics (including seismology and earthquake prediction), meteorology to solar-terrestrial physics and astronomy. Nonetheless, strong networks have been established by staff in these institutes involved in the operation of observatories, for example the biannual IAGA observatory workshops, INTERMAGNET and the World Data Centres for Geomagnetism. The ESA Swarm mission will take advantage of these networks.

Processing satellite magnetic observations often initially involves their selection on the basis of data obtained by ground observatories, usually in the form of geomagnetic indices like K_p and D_{st}. Also, many Swarm L2 products are derived using observatory hourly mean data and/or indices and some require observatory data products for their validation. Thus observatory data products are used in satel-

lite data *selection*, L2 product derivation and validation. We describe briefly how the K_p and D_{st} index data sets are prepared and then give details of the quality control measures we found were required for the recent historic hourly mean data sets. Efforts to improve the real-time availability of absolute observatory data are then discussed, before a short section with some concluding remarks.

2. Preparation of Observatory Indices

K_p is a 3-hour range index describing global geomagnetic activity (e.g. Siebert and Meyer, 1996). It is routinely derived by GFZ Helmholtz Centre in Potsdam using 1-minute data from 13 observatories. The D_{st} index monitors the strength of the axisymmetric part of the magnetospheric ring current (e.g. Sugiura and Kamei, 1991) and is determined using hourly mean values from 4 observatories by the World Data Centre (WDC) for Geomagnetism at Kyoto. Both indices are reformatted and delivered by GFZ Helmholtz Centre for distribution through the ESA Payload Data Ground Segment (PDGS). K_p and its linear equivalent ap is provided with time given in Modified Julian Day 2000 (as is the case for most Swarm AUX data products) in the auxiliary product with filename similar to `SW_OPER_AUX_KP__2__`
`YYYYMMDDThhmmss_YYYYMMDDThhmmss_VVVV.DBL` where "YYYYMMDDThhmmss" stands for time stamp of first, respectively last, data point in the file and "VVVV" is a four digit version number. Data for January 1st 1998 to January 31st 2011 are for instance available in the file `SW_OPER_AUX_KP__2__19980101T000000_`
`20110201T000000_0001.DBL`. D_{st} and its decomposition into magnetospheric (external) and internally induced parts using a 1-D model of electrical conductivity of the Earth's mantle (e.g. Maus and Weidelt 2004; Olsen *et al.* 2005) are provided as time series of 1 hour

sampling rate in the auxiliary data product with filename similar to `SW_OPER_AUX_DST_2__YYYYMMDDThhmmss_` `YYYYMMDDThhmmss_VVVV.DBL`.

3.   Preparation of Observatory Hourly Mean Data

Observatory data are extracted from data holdings at the WDC for Geomagnetism at Edinburgh. This WDC, and the Kyoto WDC, are part of the ICSU (International Council for Science) World Data Centre System and ensures the long-term availability of geomagnetic data for research. Although data quality checks have been made on incoming data over the years, and feedback provided to the data suppliers, the WDC is ultimately a depository. It therefore contains data of varying quality and use of these data must be preceded by a selection procedure. In preparation of the Swarm mission, and in order to support the analysis of the pre-Swarm satellite data (in particular Ørsted and CHAMP) we provide a consistent data set of quality-checked observatory hourly mean values for the years 1997 and onwards. When Swarm is launched these data will be augmented with quasi-definitive data mainly from INTERMAGNET observatories. In later releases of the observatory hourly mean data series these quasi-definitive data will be replaced by definitive data, once available. INTERMAGNET is an organisation involved in promoting standards in digital observatory operations around the world and has facilitated the improvement in observatory data quality since the early 1990s. More information can be found at www.intermagnet.org.

For all new data received for assimilation into the WDC holdings, quality control (QC) procedures are undertaken. A range of computer programs is used for this QC, the main one being the INTERMAGNET data viewer program *imcdview* developed by BGS. Values are then compared with a global model to check for gross errors such as incorrect signs.

Observatory hourly mean values for use in conjunction with the analysis of magnetic satellite data like those from Swarm should

- contain minimal measurement noise
- be as complete as possible (no data gaps in time)
- be corrected to absolute values over multi-year periods, i.e. drift-free
- be without discontinuities
- comply with an agreed data format.

In addition to these requirements it is helpful if the data are

- in the same coordinate frame as the satellite data
- time- and position-stamped like the satellite data.

Thus, additional QC and post-processing of the hourly mean values held at the WDC and provided by INTERMAGNET is required. This is an iterative process with a quiet-time selection of data over several years first being made and any spikes identified and removed manually. Discontinuities are then identified. Discontinuities are sometimes measured from simultaneous measurements at the two sites if they are caused by a move of the instrument pillar, or quantified by other means. In these cases the data should be processed by the institute in such a way
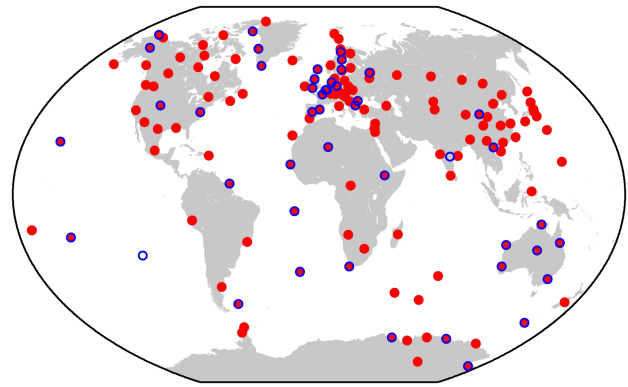

Fig. 1. Locations of observatories with usable definitive data 1997–2010. Outlined in blue are the INTERMAGNET observatories supplying quasi-definitive data in 2012.
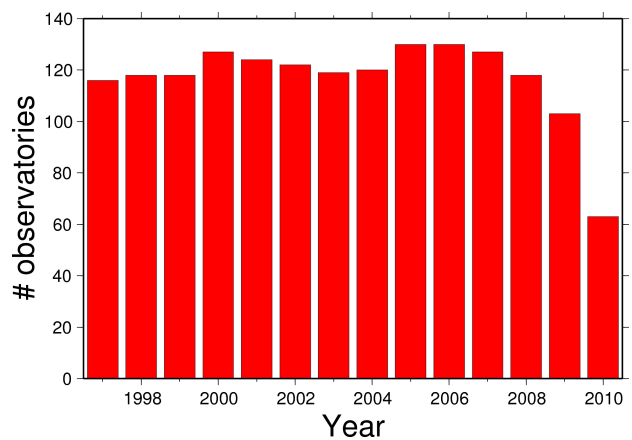

Fig. 2. Distribution in time of observatory definitive data (as of October 2011).

that the discontinuities are translated to the start of the year, and recorded in the observatory's series of annual mean values. The known discontinuity values are stored in a separate file and applied to the hourly mean values to produce a series without steps. However, very often the discontinuities are not known or cannot be quantified and in these cases the observatory data were split into two or more series. To ease this process 4-letter observatory codes were introduced comprising the 3-letter IAGA code and a numeral. The process is then repeated until all hourly mean values over the period of interest can be selected and plotted without obvious large spikes and free of all discontinuities.

In the following we illustrate the process with data spanning the years 1997 (i.e. two years before the launch of the Ørsted satellite in February 1999) to 2010. The resulting data set contains hourly mean values from 144 observatories, the spatial distribution of which are shown in Fig. 1 and the temporal distribution in Fig. 2. There are data discontinuities (jumps) in 35 observatories; in the case of 15 observatories (BDV, DOB, ESK, HER, HBK, HLP, IQA, LER, NVS, PPT, PST, SFS, TRO, TSU and WNG) these discontinuities have been known (since reported by the observatories in their annual mean values). In a few cases the discontinuities in the hourly mean series were at different

Table 1. Format description for the auxiliary Swarm product AUX_OBS_2_.

```
Variable number of header lines each starting with '#'. Note, embedded header lines expected,
such as introducing a next station (see example below)
Missing value flag = NaN (this is in keeping with other Swarm products)
```

Column	Format	Description
1-4	s4	3-letter IAGA code, followed by digit
5-12	f8.3	geocentric latitude (degrees)
13-20	f8.3	geocentric longitude (degrees)
21-29	f9.3	radius (km)
30-34	i5	Year
35-37	i3	Month
38-40	i3	Day
41-45	f5.1	UT (hours)
46-54	f9.1	North (-B_theta) component (nT)
55-63	f9.1	East (+B_phi) component (nT)
64-72	f9.1	Centred (-B_r) component (nT)

```
Example:
#Observatory hourly mean values selected from WDC Edinburgh/INTERMAGNET
#File created on   6  2 2012
#obs  gc_lat    long      rad yyyy mm dd  UT N=-Btheta    E=Bphi    C=-Br
################################################################################
#AAE1
AAE1   8.974  38.767 6380.055 1998  1  1  0.5   35961.0       NaN    968.0
AAE1   8.974  38.767 6380.055 1998  1  1  1.5   35960.0     746.0    968.0
....
AAE1   8.974  38.767 6380.055 1998 12 31 22.5   35987.0     758.0   1036.0
AAE1   8.974  38.767 6380.055 1998 12 31 23.5   35988.0     756.0   1036.0
#ABK0
ABK0  68.218  18.817 6360.064 1998  1  1  0.5   11466.0     966.0  51282.0
ABK0  68.218  18.817 6360.064 1998  1  1  1.5   11467.0     965.0  51281.0
....
```

times to those in the annual means series where the discontinuities are recorded. In addition we found jumps in the data of 20 observatories (ABG, API, BMT, BNG, BSL, CDP, GUI, HUA, IQA, KSH, LVV, LZH, NGP, PHU, QGZ, QIX, SSH, SUA, TUC and VSK). Since the sizes of these jumps are not exactly known it was necessary to split these data series (ABG1, ABG2, API1, API2 and so on).

The following 29 observatories have some data spanning one or more years during the period 1997–2010 which are deemed of insufficient quality and are not used: AAA, ABG, API, ARS, BGY, BMT, BNG, CNH, DLR, ELT, HBK, IQA, KNZ, KSH, LVV, MIZ, MOS, NGP, PHU, PST, SIL, THY, TIR, TND, TSU, VNA, VSK, VSS and WHN.

Five observatories (API, BOX, KIR, MCQ and TEO) have typographical errors in their data during the period 1997–2010 which were corrected during this project. These typographical errors are incorrect flag values and are corrected in the data available from the WDC. A note is added to this effect in the observatory metadata and the original files are retained internally by the WDC.

The following 7 observatories have some data during the

period 1997–2010 which are flagged as erroneous (generally spanning a few days, sometimes just one component): ARS, BGY, HLP, LRM, PAG, SFS and TIR.

The magnetic vector components of ground observatory data are usually provided in the geodetic frame (e.g. Z being the vertical downward component with respect to an ellipsoid) and site coordinates are usually provided in geodetic coordinates. In contrast, satellite data are typically provided in a geocentric frame (e.g. B_C is the magnetic field component pointing to the centre of the Earth) with position in geographic spherical coordinates (radius, co-latitude and longitude). Since the main purpose of the observatory data set described in this paper is use in combination with satellite data, all ground observatory data are converted from geodetic components and coordinates to geographic spherical components and coordinates. Table 1 presents an example of the final observatory data.

Data from all observatories for one year are provided in one file with filename similar to SW_OPER_AUX_OBS_2__ YYYYMMDDThhmmss_YYYYMMDDThhmmss_VVVV.DBL (data for the year 2015 will for instance be made available
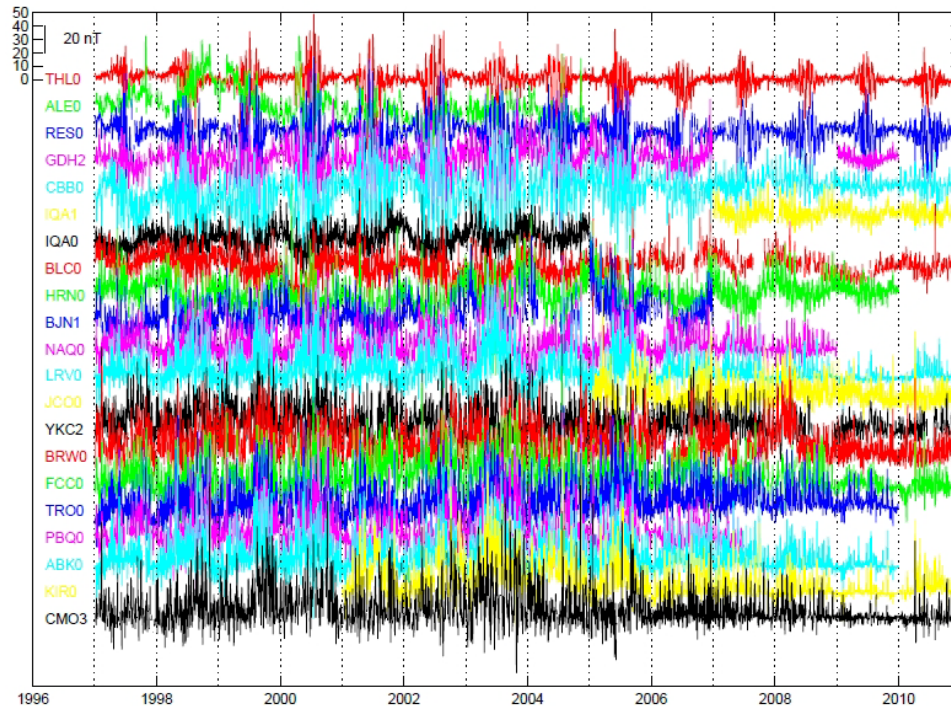
Fig. 3. Time series of residuals in the dipolar north direction for high latitude observatories.
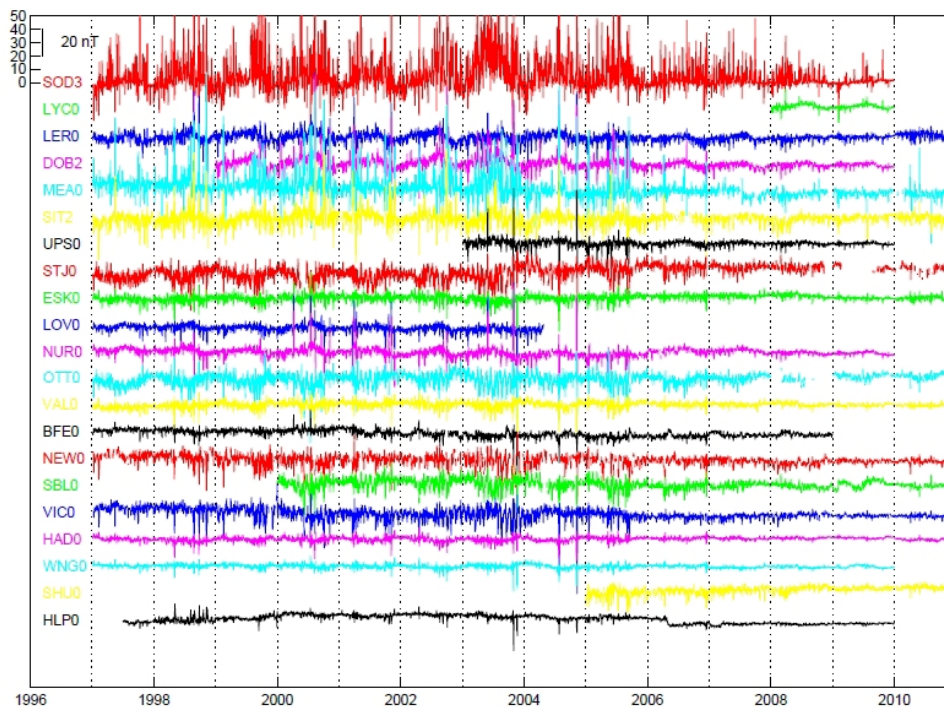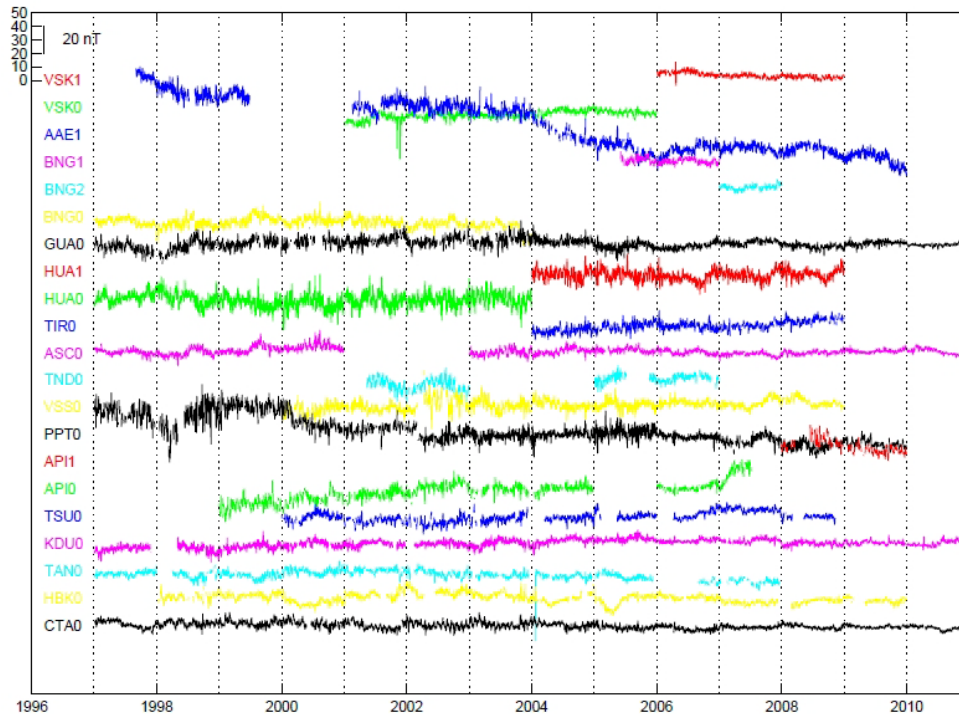

Fig. 4. Time series of residuals in the dipolar north direction for high and mid latitude observatories.

in the file `SW_OPER_AUX_OBS_2__20150101T003000_ 20151231T233000_0001.DBL`).   During the operational phase of the Swarm satellite mission the data will be updated every 3rd month and will be distributed by the ESA PDGS similar to all other Swarm data products, and also directly from ftp://ftp.nerc-murchison.ac.uk/geomag/smac/AUX_OBS_2/. The TEST

(replacing OPER in the names above) files for 1997–2010 containing the data presented in this paper are available from this ftp address.

4.   Hour-by-Hour Spherical Harmonic Analysis

As is often the case with scientific data QC, it is the detailed scientific analysis that uncovers remaining prob-

Fig. 5. Time series of residuals in the dipolar north direction for low and southern hemisphere mid latitude observatories.
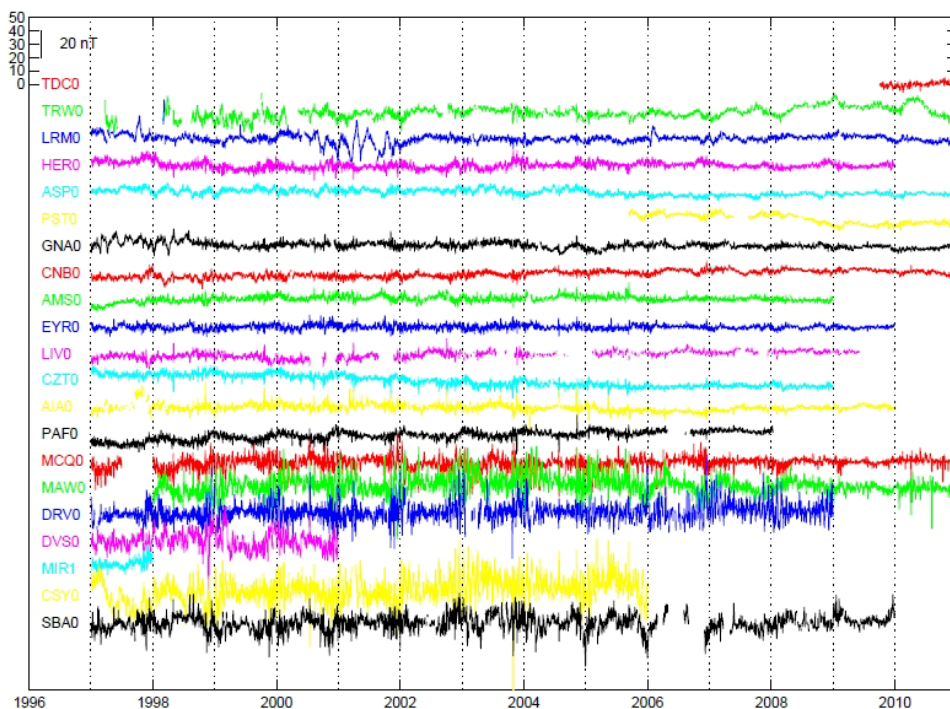
Fig. 6. Time series of residuals in the dipolar east direction for southern hemisphere mid latitude observatories.

lems in the data. This is particularly the case with geomagnetic data where it is difficult to detect small discontinuities, spikes and drifts over and above the natural magnetic field variations. These variations can be both very small e.g. annual secular variation and oceanic signals, and large, e.g. during magnetic storms. For the observatory hourly means further problems were discovered by fitting hour-by-hour

spherical harmonic models to the data with *a priori* estimates of known signals removed.

The core and large-scale crustal signal was first removed using the latest in the CHAOS series of models (Olsen *et al.*, 2010). The ionospheric (primary and induced) contributions as predicted by CM4 (Sabaka *et al.*, 2004) were then removed. The remaining crustal field was determined
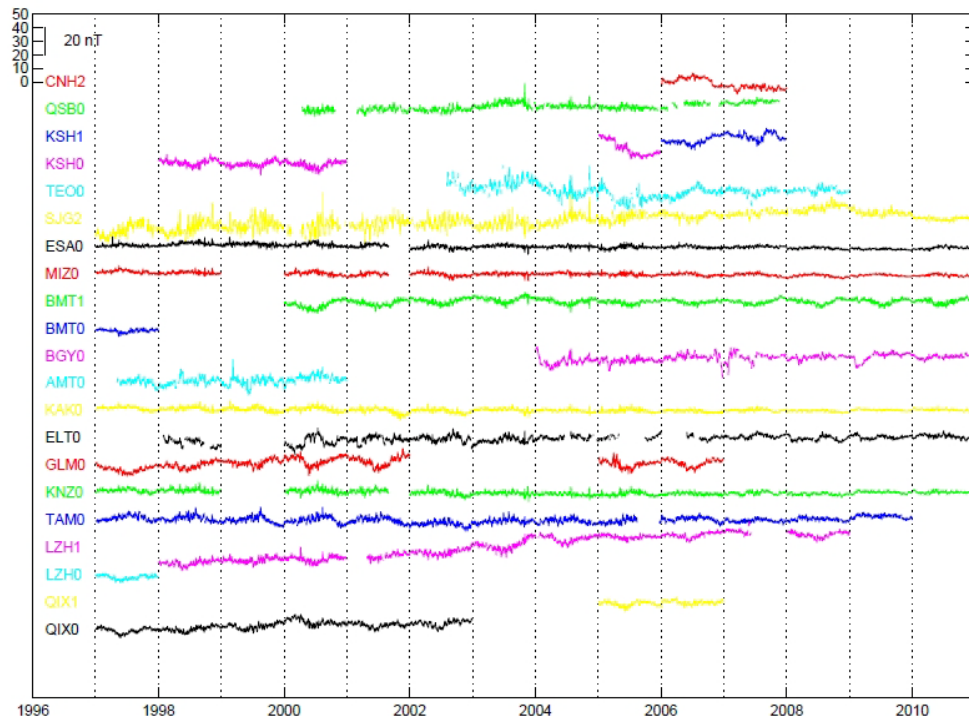
Fig. 7. Time series of residuals in the vertical direction for mid latitude observatories.

and removed by subtracting the mean of the local night-time values during geomagnetic quiet periods ($K_p < 2+$, $|dD_{st}/dt| < 2$ nT/hr). The data were then rotated into a dipole coordinate system and hour-by-hour robust spherical harmonic models up to degree 9 and order 1 were fitted to the horizontal and vertical data separately. This procedure does not produce realistic magnetospheric field models but it does help reduce any potential and non-potential signals that are coherent in time and space in particular regarding neighbouring sites. The time series of the residuals to these models are then plotted in order of dipole latitude.

Figures 3–5 show a selection of the observatory residual time series in the dipolar north direction, Fig. 6 in the dipolar east direction and Fig. 7 in the vertical direction. Figure 3 illustrates that a significant signal remains in the residuals at polar latitudes. This is expected since a global spherical harmonic model of degree 9 is not able to describe the spatial scales of the magnetic signature of the auroral electrojets. (Describing those would require spherical harmonics at least up to degree 20 or so, but estimating such a model is hardly possible with data from the spatially very uneven observatory network.) The coherency in the residuals between observatories at similar dipole latitudes (and longitudes) confirms that the source of the contamination is natural, most likely dominated by the ionosphere. An example of this coherent signal, probably from the auroral electrojet, can be seen in LER0 (Lerwick in northern UK) and DOB2 (Dombås in mid-Norway) in Fig. 4. These coherent signals in the residuals decrease in magnitude as the dipole latitude decreases since the model is able to describe the more large-scale structure of external variations at non-polar latitudes and data errors become visible. Examples of data problems in Fig. 4 include small steps in the data

in SBL0 (Sable Island, offshore Canada) during 2009 and a small step/drift in HLP0 (Hel, Poland) during 2006.

Figure 5 shows residuals from observatories near the dipole equator and at low dipole latitudes in the southern hemisphere. The many gaps in Fig. 5 demonstrate that a considerable amount of data has already been rejected, or are missing. Remaining indicators of data problems include a curious drift over several years in AAE1 (Addis Ababa, Ethiopia), a spike in VSK1 (Visakhapatnam, India) in 2006, noise in PPT0 (Pamatai, French Polynesia) during 1997–99, drift/step in PPT0 during 2000 and 2002, drift in API0 (Apia, Western Samoa) during 2007 and a drift in HBK0 (Hartebeesthoek, South Africa) during 2005.

Figure 6 is an example of the residuals in the dipolar east direction at mid to high dipole latitudes in the southern hemisphere. An example of a data problem is seen in the LRM0 (Learmonth, Australia) residuals for 2000–2001. The residuals in the vertical direction (Fig. 7) highlight further data problems, for example the drift in LZH1 data (Lanzhou, China). In contrast, the residuals for the Japanese observatories in this plot (ESA0–Esashi, MIZ0–Mizusawa, KAK0–Kakioka and KNZ0–Kanozan) indicate excellent data quality. The residuals in the vertical direction generally have larger amplitudes, especially at high dipolar latitudes, due to the larger unmodelled fields and induction effects for this component.

These examples of problems in the data and visible in Figs. 3–7 are not always easy to deal with, even by the institutes operating the observatories and responsible for processing their data. There are many factors affecting quality of observatory data (see for example chapter 6, St-Louis, 2011) but the following are considered to be the more difficult ones to deal with and which might manifest themselves

Table 2. INTERMAGNET observatories with quasi-definitive data for months in 2012 (as of 30/4/2013).

```
        JFMAMJJASOND            JFMAMJJASOND

 AAE  xxxxxxxxxxxx      KOU  xxxxxxxxxxxx
 ABK  xxxxxxxxxxxx      LER  xxxxxxxxxxxx
 ASC  xxxxxxxxxxxx      LRM  xxxxxxxxxx--
 ASP  xxxxxxxxxxxx      LYC  xxxxxxxxxxxx
 BOU  xxxxxxxxxxxx      LZH  --xxxxxxxxxx
 BOX  xxxxxxxxxxxx      MAB  xxxxxxxxxx-x
 BRW  ----------xx      MAW  xxxxxxxxxxx-
 CLF  xxxxxxxxxxxx      MBO  xxxxxxxxxxxx
 CMO  xxxxxxxxxxxx      MCQ  xxxxxxxxxxx-
 CNB  xxxxxxxxxxxx      NAQ  --xxxxxxxxxx
 CSY  xxxxxxxxxxxx      NGK  xxxxxxxxxxxx
 CTA  xxxxxxxxxxxx      PAG  xxxxxxxxxxxx
 DOU  xxxxxxxxxx-x      PHU  xxxxxx----xx
 EBR  ------xxxxxx      PPT  xxxxxxxxxxxx
 ESK  xxxxxxxxxxxx      PST  xxxxxxxxxxxx
 FRD  xxxxxxxxxxxx      SBA  ----------x
 GDH  xxxxxxxxxxxx      SPT  xxxxxxxxxxxx
 GNG  xxxxxxxxxxxx      SUA  xxxxxxxxx---
 HAD  xxxxxxxxxxxx      TAM  xxxxxxxxxxxx
 HER  -------xxxx       TDC  xxxxxxxxxxxx
 HON  xxxxxxxxxxxx      THL  --------xxx
 HYB  xxxxxxxxxxxx      UPS  xxxxxxxxxxxx
 IPM  xxxxxxxxxxxx      WNG  xxxxxxxxxxxx
 KDU  xxxxxxxxx---
```

in series of hourly means: stability of measurement platforms, changing temperature coefficients, changing site differences, observer skill in absolute observations, presence of bias in absolute observations and high spatial gradients in the magnetic field in vicinity of sensors. Some problems remain in the data as they are deemed not to be significant enough to render them useless for all scientific studies. For example, long-term drifts and steps are not considered to devalue the data significantly for studies of external field variations. Likewise, noise and spikes do not have have a big effect on long-term secular variation studies.

5. Quasi-Definitive Data

One difficulty in using observatory data for producing models of the Earth's magnetic field is that there is often a long delay till the final definitive data are released. The reasons for this vary from one institute to another and from one observatory location to another and may be related to the presence of the afore-mentioned data quality issues, lack of man-power and the tradition of annual processing for the production of yearbooks. Often, by the time the definitive data are released, they have passed their period of potential peak usefulness. In contrast, satellite data are often available after a few days or weeks after acquisition. A time delay in the availability of observatory data therefore hampers the optimal use of these data in combination with most recent satellite observations.

Fortunately INTERMAGNET has taken the initiative to try and improve this situation by developing a new type of data, called quasi-definitive data (Peltier and Chulliat, 2010). These data are intended to be within 5 nT of the final definitive data and are required to be submitted within 3 months of collection. This initiative started in late 2011 and by April 2013 the 47 observatories listed in Table 2 (and also plotted in Fig. 1) are submitting quasi-definitive data to INTERMAGNET. There are currently over 100 INTERMAGNET observatories so there is still much work to be done. Note also that for a few of the 47 observatories there appears to be data flow problems.

6. Concluding Remarks

In this paper we have concentrated on hourly mean data from observatories because the amount of data that have to be processed and analysed simultaneously with satellite data is easier to handle compared e.g. to using one-minute data. We have described the additional quality control measures we found necessary for the observatory hourly means spanning the current magnetic survey satellite era, the format of the files and where online they will be made available with a 3-month update cycle.

It should be noted that the standard product from observatories is one-minute mean data, with work underway at many observatories to move to one-second data. One-minute data are available directly from the World Data Centres in Edinburgh (www.wdc.bgs.ac.uk), Kyoto (wdc.kugi.kyoto-u.ac.jp), Moscow (www.wdcb.ru) and Boulder (spidr.ngdc.noaa.gov) and from INTERMAGNET (www.intermagnet.org). The problems we have found in the hourly mean values (which are based on minute values) are very likely also present in data of higher temporal resolution, and once hourly mean values have been flagged as erroneous or data jumps and spikes have been identified it

should be possible to correct the one-minute data.

One-second data are becoming more widely available now with much development in sensors, digitisers and in data distribution. With the Swarm satellite instruments sampling the field at 1 Hz and higher, and with many interesting high frequency signals to be analysed and understood, it seems likely that these higher cadence observatory products will be very useful.

References

Maus, S. and P. Weidelt, Separating the magnetospheric disturbance magnetic field into external and transient internal contributions using a 1D conductivity model of the Earth, *Geophys. Res. Lett.*, **31**(12), L12614, doi:10.1029/2004GL020232, 2004.

Olsen, N., T. J. Sabaka, and F. Lowes, New parameterization of external and induced fields in geomagnetic field modeling, and a candidate model for IGRF 2005, *Earth Planets Space*, **57**, 1141–1149, 2005.

Olsen, N., M. Mandea, T. J. Sabaka, and L. Tøffner-Clausen, The CHAOS-3 geomagnetic field model and candidates for the 11th Generation IGRF, *Earth Planets Space*, **62**, 719–727, 2010.

Peltier, A. and A. Chulliat, On the feasibility of promptly producing quasi-definitive magnetic observatory data, *Earth Planets Space*, **62**(2), e5–e8, doi:10.5047/eps.2010.02.002, 2010.

Sabaka, T. J., N. Olsen, and M. E. Purucker, Extending Comprehensive Models of the Earth's magnetic field with Ørsted and CHAMP, *Geophys. J. Int.*, **159**(2), 521–547, 2004.

Siebert, M. and J. Meyer, Geomagnetic Activity Indices, in *The Upper Atmosphere*, edited by W. Dieminger, G. K. Hartmann, and R. Leitinger, pp. 887–911, Springer, 1996.

St-Louis, B. J. (ed), INTERMAGNET Technical Reference Manual, version 4.5, INTERMAGNET, http://www.intermagnet.org/publications/im_manual.pdf (accessed 30-04-13), 2011.

Sugiura, M. and T. Kamei, Equatorial Dst index, 1957–1986, in *IAGA Bulletin 40*, edited by Berthelier, A. and M. Menvielle, ISGI Publ. Off. Saint Maur, 1991.

S. Macmillan (e-mail: smac@bgs.ac.uk) and N. Olsen