

Resolving issues concerning Eskdalemuir geomagnetic hourly values

S. Macmillan and E. Clarke

British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, UK

Received: 24 January 2011 – Revised: 31 January 2011 – Accepted: 1 February 2011 – Published: 8 February 2011

Abstract. The hourly values of the geomagnetic field from 1911 to 1931 derived from measurements made at Eskdalemuir observatory in the UK, and available online from the World Data Centre for Geomagnetism at <http://www.wdc.bgs.ac.uk/>, have now been corrected. Previously they were 2-point averaged and transformed from the original north, east and vertical down values in the tables in the observatory yearbooks. This paper documents the course of events from discovering the post-processing done to the data to the final resolution of the problem. As it was through the development of a new index, the Inter-Hour Variability index, that this post-processing came to light, we provide a revised series of this index for Eskdalemuir and compare it with that from another European observatory. Conclusions of studies concerning long-term magnetic field variability and inferred solar variability, whilst not necessarily consistent with one another, are not obviously invalidated by the incorrect hourly values from Eskdalemuir. This series of events illustrates the challenges that lie ahead in removing any remaining errors and inconsistencies in the data holdings of different World Data Centres.

Keywords. Geomagnetism and paleomagnetism (Time variations, diurnal to secular; Time variations, secular and long term; General or miscellaneous)

1 Introduction

For the past few years it has been known that the Eskdalemuir hourly values from 1911 to 1931 were post-processed with unfortunate consequences. These data were formerly available online from the World Data Centre (WDC) for Geomagnetism at the Danish Meteorological Institute till the start of 2007 and then at the authors' institute. The post-processing was the application of a 2-point running mean and

transformation to declination, horizontal and vertical intensities (DHZ) from the original northerly, easterly and vertical intensities (XYZ) recorded in the tables of the observatory yearbooks. The reason for the post-processing was presumably to make these earlier hourly values compatible with the later values which were hourly mean DHZ values centred on the half hour (Eskdalemuir yearbooks, 1932 onwards). There is no record of when this post-processing was done. The 1911–1917 data were XYZ spot values on the hour (according to yearbook table headings but more on this later) and the 1918–1931 data were XYZ mean values centred on the hour (Eskdalemuir yearbooks, 1911–1931). This post-processing had the effect of lowering the spectral content for high frequency studies performed on the hourly value time series. However the significance of this only really came to light with the development of the Inter-Hour Variability index in the 2000s.

2 The discovery

The first indication of something unexpected in the early Eskdalemuir hourly values was from correspondence with Leif Svalgaard in 2004 where he presented ratios of the newly defined Inter-Hour Variability indices for the UK observatories. The IHV index is the sum of absolute differences of hourly means (or values) for a geomagnetic component from one hour to the next over the six hour interval around local midnight. The ratios for the annual means of the IHV indices for the horizontal component as recorded at the UK observatories at Eskdalemuir, Lerwick, Abinger and Hartland revealed a sharp change in the Eskdalemuir record at the end of 1931 (Fig. 1).

The creation of the IHV index was motivated by issues concerning the *aa* index (Svalgaard et al., 2003). The *aa* index has been critical for inferring long-term changes in the Sun (e.g. Lockwood et al., 1999) as it is the longest running index of magnetic activity available, dating back to 1868. However some records of observatory hourly values are also very long and this lead to the development and



Correspondence to: S. Macmillan
(smac@bgs.ac.uk)

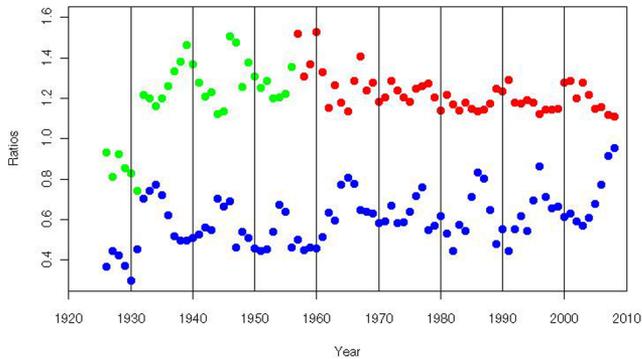


Fig. 1. Ratios of annual means of UK observatory IHV indices based on data available at the Edinburgh WDC before 9 August 2010. Green is ratio of Eskdalemuir to Abinger, red is ratio of Eskdalemuir to Hartland and blue is ratio of Eskdalemuir to Lerwick.

analyses of IHV indices for inferring long-term changes in the Sun and near-Earth space (Svalgaard et al., 2004; Mursula et al., 2004; Mursula and Martini, 2006; Svalgaard and Cliver, 2007a, b). The problem with the Eskdalemuir IHV series was known at this stage and attempts were made to correct it (Martini and Mursula, 2006). Another point to note at this stage was that the World Centres were acknowledged collectively for the supply of the hourly values and the individual WDC supplying the data was not identified. This turns out to be important. In the meantime the long-term solution to the problem seemed to lie with either undoing the post-processing or redigitising the original hourly value tables in the observatory yearbooks.

3 Undoing the post-processing

In theory, it should be possible to undo such relatively simple post-processing by obtaining an initial value from the yearbooks and applying it to each series of post-processed data using

$$E_{i+1} = 2E'_i - E_i$$

where E_i is an original hourly value of an element of the Earth's magnetic field for hour i and E'_i is a 2-point averaged hourly value of an element of the Earth's magnetic field for hour i . Knowledge of E_0 (the value at the start of the time series) should be all that is required to regenerate the original series of data. However it turns out that the transforming from XYZ to DHZ and the limitations of the WDC format significantly complicates matters. We explain this now in more detail.

The digital storage and exchange format that is widely used for hourly values is the WDC format, described at <http://www.wdc.bgs.ac.uk/catalog/master.html>. This is the natural extension of the way that hourly values were presented in the older yearbooks. The use of a baseline value,

one line of data per element per day and 1 nT resolution is common to both the yearbook tables and the WDC-format files. The unit for declination of the magnetic field in the WDC-format file is tenth arc-minute, and the resolution is 1 tenth arc-minute.

There are 4 options of how the post-processing was done:

1. The XYZ values were transformed to DHZ, then 2-point averaged and rounded to 1 nT.
2. The XYZ values were transformed to DHZ, rounded to nearest 1 tenth arc-minute and 1 nT (the resolution of the WDC format), then 2-point averaged and rounded to 1 nT. This option is possible if WDC-format files of unaveraged DHZ existed at an intermediate stage.
3. The XYZ values were 2-point averaged, then transformed to DHZ and rounded to nearest 1 tenth arc-minute and 1 nT.
4. The XYZ values were 2-point averaged, rounded to nearest 1 nT, then transformed to DHZ and rounded (again) to nearest 1 tenth arc-minute and 1 nT. This option is possible if WDC-format files of averaged XYZ existed at an intermediate stage.

Within each of these options are choices for rounding 0.5 to the resolution of the WDC-format files which is 1. Whilst the default rounding in the IEEE 754-2008 floating point arithmetic standard is to round to even, its implementation is dependent on computer platform and programming language. Rounding to zero and rounding up are other possibilities. Also unknown is whether the same rounding was applied to each component. A further complication is that the Y component in the yearbooks assumes west is positive, but it is not known whether this was the case in any intermediate WDC-format files.

Because of these uncertainties in the post-processing, and also because any resulting values from attempts to undo the post-processing would require comparison with the original data sources, the yearbook tables were scanned and digitised. It should be noted that the original data are the analogue magnetograms and tables of absolute observations and whilst there is a programme underway in BGS to digitise these for long-term preservation, it was not deemed practical to use these to obtain new series of absolute hourly values within a reasonable timescale.

4 Experiences with scanning and digitising yearbook tables

In support of a study by Martini and Mursula (2006), work began to redigitise the original hourly values directly from the monthly tables in the Eskdalemuir observatory yearbooks. Originally this involved making scanned copies of only the January tables for each year as described in Martini

TERRESTRIAL MAGNETIC FORCE : WEST COMPONENT.
Mean values for periods of sixty minutes centred at the Hours of Greenwich Mean Time.

285. Eskdalemuir. (—Y.) 4,000 γ (0.4 C.G.S. unit) — **June, 1926.**

Hour. G.M.T.	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	Noon.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	Mean.
Day.	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ	γ
1 D	486	480	479	476	471	459	453	451	445	447	491	502	512	512	557	506	541	551	539	531	506	472	465	422	430	488
2 D	420	313	332	373	405	452	420	459	453	492	488	491	504	512	519	521	519	516	520	499	492	481	479	480	484	468
3	484	497	470	458	464	453	437	431	432	437	445	405	480	495	498	497	495	494	485	485	485	485	476	478	467	472
4 Q	467	465	465	465	463	454	449	445	449	445	405	479	496	505	512	510	502	496	489	484	484	485	474	474	473	476
5	473	466	471	467	467	459	451	445	443	451	465	474	489	498	505	511	500	492	492	490	488	478	483	480	482	477
6	482	478	474	471	465	452	439	436	439	445	459	479	497	512	512	511	502	498	500	502	493	453	478	484	484	478
7	481	472	452	453	445	447	442	433	440	447	470	487	505	512	510	512	521	505	498	496	495	503	499	492	484	480
8 D	484	492	443	465	474	484	465	437	439	439	470	494	504	510	525	516	515	493	491	485	491	478	491	459	448	480
9 D	447	450	443	440	459	448	454	431	432	444	471	489	504	511	511	512	516	497	498	493	494	471	486	469	469	474
10	469	458	458	458	460	442	437	444	448	451	458	471	483	499	496	504	505	511	500	498	492	491	484	479	477	475
11	477	472	452	446	452	457	446	437	434	437	452	473	492	504	512	515	506	498	490	491	490	491	483	478	481	474
12 Q	481	477	477	471	465	457	450	446	445	450	457	477	497	512	524	516	511	497	492	497	491	497	486	485	483	481
13	483	479	477	473	468	464	451	456	457	458	466	487	497	514	522	528	519	506	497	499	490	479	466	466	484	480
14	466	466	464	460	465	462	452	448	445	451	459	471	486	505	516	514	510	505	497	495	486	484	484	483	477	478
15	477	467	466	464	466	465	451	450	452	448	475	479	490	499	500	504	504	505	507	500	496	490	477	485	483	480
16	483	480	492	499	477	451	450	448	442	441	464	476	493	499	497	505	505	498	487	496	494	493	482	485	481	484
17	487	477	477	481	479	458	451	452	452	471	471	479	490	497	504	505	504	502	490	493	480	485	484	480	477	482
18	477	476	483	483	490	467	457	448	448	451	463	471	483	497	498	498	496	501	490	489	480	491	483	479	472	470
19	477	477	471	478	483	465	459	451	450	451	471	484	484	489	497	497	497	493	495	491	488	485	483	481	479	479
20 Q	481	477	477	470	466	457	448	442	444	452	459	470	491	499	499	497	497	496	497	485	485	483	481	481	483	477
21	483	477	477	470	463	453	451	446	451	458	475	483	495	497	502	499	492	492	495	491	491	487	484	484	483	470
22	483	479	478	479	476	461	454	448	448	458	477	495	503	505	504	498	497	491	488	483	483	484	483	490	491	481
23 D	491	491	481	481	464	452	445	452	451	465	483	502	502	522	524	543	528	524	507	484	476	481	480	483	486	485
24	466	469	477	471	471	453	458	463	457	459	471	487	487	507	504	512	505	496	508	470	478	477	480	477	476	478
25 Q	476	475	472	471	467	458	452	444	444	443	454	468	480	504	511	516	506	500	500	485	484	483	481	479	477	477
26 Q	477	476	476	471	468	459	446	439	440	444	460	477	492	497	497	503	503	503	503	493	483	483	483	484	484	486
27	484	483	483	476	469	460	451	446	436	439	452	471	491	510	521	524	524	527	507	493	490	485	485	483	481	484
28	487	471	468	466	463	450	446	448	445	446	451	466	491	511	511	524	524	524	507	497	491	484	479	474	481	470
29	481	471	474	470	461	445	431	431	437	459	464	476	483	490	503	510	510	503	492	485	476	477	477	474	474	476
30	474	471	470	469	464	460	457	450	452	462	483	496	496	498	503	503	503	503	491	490	488	488	488	474	475	478
Mean.	476	469	466	466	467	457	449	445	445	451	463	478	493	504	510	512	506	503	493	489	485	484	477	485	485	488

Q denotes an "International Quiet Day," while D denotes a disturbed day used for the computation of Tables 318-329.

Fig. 2. Example of a TIFF file of a monthly table of hourly values scanned from the ESK yearbooks. This illustrates the typical poor quality of the original tables.

and Mursula (2006), however this work continued to complete the full set of digital images for all months.

The yearbooks were located and transported to Edinburgh – some from the BGS archive at Hartland observatory and others from Eskdalemuir. Following tests on and selection of scanning hardware and Optical Character Recognition (OCR) software, the monthly tables for each component for 23 years (1932 and 1933 were also done, resulting in a total of over 800 tables) were scanned and saved as tagged image file format (TIFF) files. Some of the yearbooks were in degraded condition and thus the clarity of the tabulated data within the images obtained was often poor (Fig. 2). Using OCR software, attempts were made to extract the data from the TIFF files. However too many errors or unrecognised characters were obtained for each image and the process was abandoned as one that was uneconomical in time and effort. The OCR method was confirmed as too much of a challenge in this case, when example files were sent to three different specialist companies. They all recommended direct data entry rather than OCR. One company was commissioned to carry out the data entry work.

The digitised XYZ data were returned in Microsoft Excel spreadsheet format, with each table on a single worksheet as

per its original layout in the yearbook. The tables included the daily averages as well as the averages over each UT hour in the month. These averages would prove useful for quality control (QC) purposes. The data were transformed into DHZ and two-point averaging applied. This enabled comparison with the existing post-processed WDC data. It was clear at this stage that there were many errors in the digitised data (e.g. Fig. 3 top and middle panels) and that the data entry company had not carried out the QC that had been assured.

Further work was then required to identify and correct the data entry errors. Firstly the Excel spreadsheets could be used to compute daily and monthly UT averages, and compared against the corresponding values entered in the spreadsheets. Any errors identified at this stage were corrected. Secondly these revised data were then compared again with the post-processed WDC data, as in the previous paragraph, and all differences greater than 10 nT identified and corrected: a task that proved to be very labour intensive. Although a large number of smaller errors remained (e.g. Fig. 3 bottom panel) it was clear that the Eskdalemuir hourly values from this procedure were a better match to the original data in the yearbooks than the post-processed data available from the WDC. These data were not released however as it

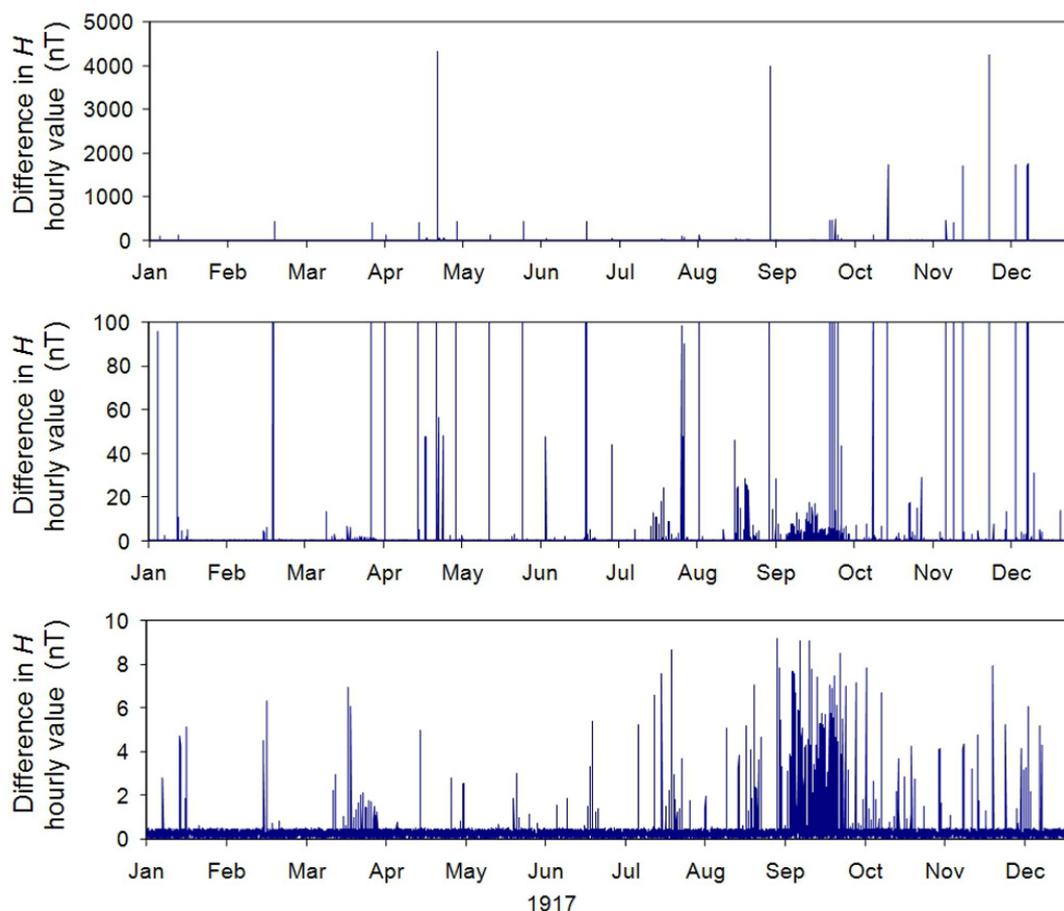


Fig. 3. Errors in the hourly mean values for 1917 from manual data entry carried out in 2007. The top panel shows full range of errors, the middle panel reduces the range to 100 nT to show more detail and the bottom panel shows the remaining errors following the corrections described in the text.

was the intention to complete the processing by removing all the remaining errors. This task has now been superseded by the discoveries and work carried out in the remainder of this paper.

5 Comparison of WDC holdings

Until recently it had been widely assumed that the WDC-format files for Eskdalemuir held at different WDCs contained the same data. However it was noted in Love et al. (2010) that the files held at Edinburgh and Kyoto WDCs were different and that the Kyoto files appeared to contain unaveraged XYZ values. A detailed comparison of the files was made by the authors, including monthly plots of differences, using the 4 options of post-processing listed in Sect. 3.

It was found that option 1 with rounding up for 0.5, resulted in the least differences between the 2 data series, particularly for Z . Table 1 lists the annual means and standard deviations of the differences in DHZ for this option. Although Table 1 shows that the overall differences are not al-

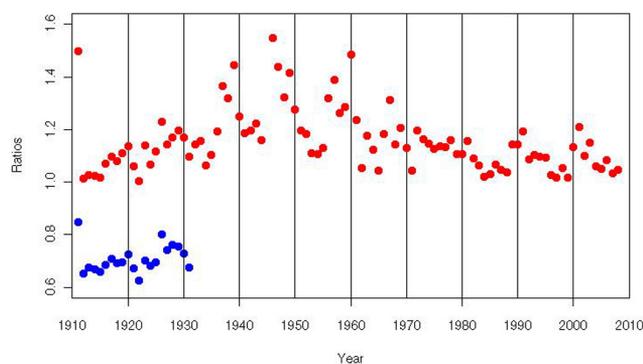


Fig. 4. Ratios of annual means of Eskdalemuir and Niemegk IHV indices based on data available at the Edinburgh WDC before (blue) and after (red) 9 August 2010.

ways zero with this option, this, along with the monthly plots of differences, was strong support for Love's claim that the data in the Kyoto WDC-format files were not post-processed.

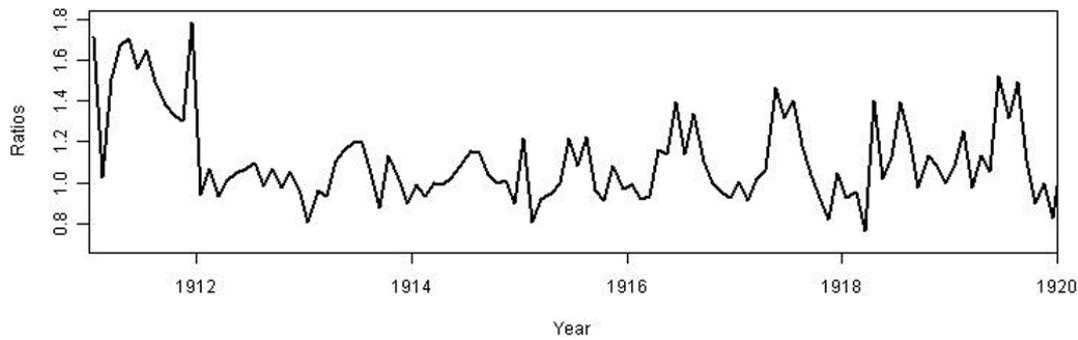


Fig. 5. Ratios of monthly means of Eskdalemuir and Niemegk IHV indices based on data available at the Edinburgh WDC after 9 August 2010.

Table 1. Annual statistics of differences between post-processed Kyoto WDC hourly values (option 1 in text) and Edinburgh WDC hourly values prior to 9 August 2010.

Year	Mean of differences in D (0.1 arc-minutes)	Standard deviation of differences in D (0.1 arc-minutes)	Mean of differences in H (nT)	Standard deviation of differences in H (nT)	Mean of differences in Z (nT)	Standard deviation of differences in Z (nT)
1911	0.0	0.1	0.0	0.1	0.0	0.0
1912	0.0	0.1	0.0	0.1	–	–
1913	0.0	0.1	0.0	0.1	0.0	0.0
1914	0.0	0.1	0.0	0.1	0.0	0.0
1915	0.0	0.1	0.0	0.1	0.0	0.0
1916	0.0	0.1	0.0	0.1	0.0	0.0
1917	1.8	87.4	–0.2	9.8	0.0	0.0
1918	0.0	0.1	0.0	0.1	0.0	0.0
1919	0.0	0.1	0.0	0.1	0.0	0.0
1920	0.0	0.1	0.0	0.1	0.0	0.0
1921	0.0	0.1	0.0	0.1	0.0	0.0
1922	0.0	0.1	0.0	0.1	0.0	0.0
1923	0.0	0.1	0.0	0.1	0.0	0.0
1924	0.0	0.1	0.0	0.1	0.0	0.0
1925	2.8	51.9	–0.4	6.8	0.0	0.0
1926	0.0	0.1	0.0	0.1	0.0	0.0
1927	0.0	0.1	0.0	0.1	0.0	0.0
1928	0.0	0.1	0.0	0.1	0.0	0.0
1929	0.0	0.1	0.0	0.1	0.0	0.0
1930	0.0	0.1	0.0	0.1	0.0	0.0
1931	0.0	0.1	0.0	0.1	0.0	0.0

Before replacing the Edinburgh files with those from Kyoto the cause of the large differences in 1917 and 1925 (Table 1) had to be found. For this the monthly plots of differences, yearbook scans and digitised values were used. For 1917 it was the July values in Y and Z in the Kyoto files that were wrong. For 1925 it was the January values in Y in the Kyoto files that were wrong and the December values in Y and Z in the Edinburgh files that contained additional errors over and above the post-processing.

Further checking of the Kyoto files revealed missing chunks of data in 1915 and 1921. Again the yearbook scans

and digitised values were used to complete the files. A final quality-control measure was to select one day at random for each year from 1911 to 1931 and to check that the values in the WDC-format files agreed with those from the yearbooks. The corrected Kyoto files were transmitted back to Kyoto and the post-processed Edinburgh files available online were replaced with the correct files on 9 August 2010.

6 Revised IHV indices

As further evidence of the reliability of the updated files, we compute IHV indices for a number of observatories and compare ratios of their annual means. In Fig. 4 it can be seen that the revised Eskdalemuir IHV index series no longer shows a large step at end of 1931 when compared with that from Niemegk. In order to investigate the anomalous value for 1911 we “zoom in” and use ratios of monthly means of IHV indices. In Fig. 5 we can see there is a step at the end of 1911. This is when the hourly values at Eskdalemuir changed from being spot values to mean values (as noted by Martini and Mursula, 2006, the headings of the yearbook tables of hourly values are not clear in this respect, but a note in the text of the yearbook for 1912 spotted by Love et al., 2010, states that the values are mean values centred on the hour). Changing from spot to mean values also happens in other observatory series (in Niemegk it happened before the start of the Eskdalemuir series) and whilst an adjustment to the resulting IHV series can be estimated for each observatory (Svalgaard and Cliver, 2007a), the ultimate remedy may lie in making high-quality scans of the original magnetograms and yearbooks and producing hourly mean values where previously only hourly spot values are available. A programme to do this is currently underway for the UK observatories. In the meantime extra information has been added to the metadata for Eskdalemuir held at <http://www.wdc.bgs.ac.uk/>. This states that 1911 hourly values are spot values centred on the hour and that 1912–1931 are mean values centred on the hour.

7 Conclusions

The significance of this work is that Eskdalemuir data are widely used in studies concerning long-term trends in magnetic field variability and inferred solar variability (for example Mursula et al., 2004; Svalgaard and Cliver, 2007b). For that reason inconsistencies and errors in the Eskdalemuir time series need to be identified and carefully corrected. Fortunately, because Eskdalemuir is but one of several observatories used in these studies, the conclusions on solar variability are unlikely to need reconsidering. However it was the detailed analyses undertaken in these solar variability studies that brought the problem reported here to light. For this we are grateful and we believe that future research using long series of hourly mean values can now be made with high confidence following the extensive analysis described here. Work is underway to improve the metadata, including time-dependent metadata, held for each observatory at the World Data Centres.

Acknowledgements. The authors thank Andrew Snedden for his help in scanning and carrying out OCR tests on the hourly value tables in the Eskdalemuir yearbooks. We also thank Kyoto WDC for supplying their Eskdalemuir WDC hourly value files. This paper is published with the permission of the Executive Director of the British Geological Survey (Natural Environment Research Council).

Topical Editor R. Nakamura thanks one anonymous referee for her/his help in evaluating this paper.

References

- Lockwood, M., Stamper, R., and Wild, M. N.: A doubling of the Sun's coronal magnetic field during the past 100 years, *Nature*, 399, 437–439, 1999.
- Love, J. J., Tsai, V. C., and Gannon, J. L.: Averaging and sampling for magnetic-observatory hourly data, *Ann. Geophys.*, 28, 2079–2096, doi:10.5194/angeo-28-2079-2010, 2010.
- Martini, D. and Mursula, K.: Correcting the geomagnetic IHV index of the Eskdalemuir observatory, *Ann. Geophys.*, 24, 3411–3419, doi:10.5194/angeo-24-3411-2006, 2006.
- Mursula, K. and Martini, D.: Centennial increase in geomagnetic activity: latitudinal differences and global estimates, *J. Geophys. Res.*, 111, A08209, doi:10.1029/2005JA011549, 2006.
- Mursula, K., Martini, D., and Karinen, A.: Did open solar magnetic field increase during the last 100 years? A reanalysis of geomagnetic activity, *Solar Phys.*, 224, 85–94, 2004.
- Svalgaard, L. and Cliver, E. W.: Interhourly variability index of geomagnetic activity and its use in deriving the long-term variation of solar wind speed. *J. Geophys. Res.*, 112, A10111, doi:10.1029/2007JA012437, 2007a.
- Svalgaard, L. and Cliver, E. W.: Long-term geomagnetic indices and their use in inferring solar wind parameters in the past, *Adv. Space Res.*, 40, 1112–1120, 2007b.
- Svalgaard, L., Cliver, E. W., and Le Sager, P.: Determination of interplanetary field strength, solar wind speed, and UV irradiance, 1890–present, in: *ESA-SP-535*, edited by: Wilson, A., ESTEC, The Netherlands, 15–25, 2003.
- Svalgaard, L., Cliver, E. W., and Le Sager, P.: IHV: a new long-term geomagnetic index, *Adv. Space Res.*, 34, 436–439, 2004.