

THE DIGITISATION OF OBSERVATORY MAGNETOGRAMS

Ellen Clarke (ecla@bgs.ac.uk), Simon Flower, Thomas Humphries, John Hussey, Robert McIntosh, Fergus MacTaggart, Brian McIntyre, Nicola Owenson, Keith Henderson, Elizabeth Mann, Kenneth MacKenzie, Simon Piper, Louise Wilson, Richard Gillanders and Karen Kilpatrick

British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom

Introduction

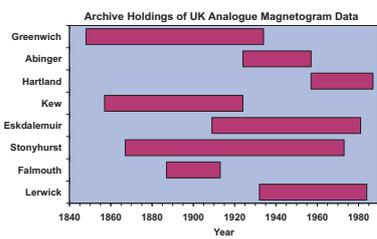
Long-running magnetic observatories are extremely important for studies of the Earth's magnetic field. They are used to model and study the core field and its secular variation and are important for analyses of the solar driven regular and irregular short-term variations in the geomagnetic field. For investigations into changes to the character of these variations over decades to centuries, data sets of most value are those that can be derived from the longest-running series of observatories.

Magnetograms from UK observatories dating back to 1848 are held in the BGS archives in Edinburgh and Hartland. In 2008 a programme of work was started to digitise the complete set of records and this work on-going. The three main aims are :

- to provide a complete backup of the collection as digital images;
- to make these images available to the scientific community on-line; and
- to develop a semi-automated method to extract digital data from the images.

We describe the method adopted to capture the magnetogram images showing some examples from the first set to be processed - Kew Observatory. We also describe the first version of a computer program specifically created to extract digital data from the images.

The Archive



Browsing the magnetogram archive, Murchison House, Edinburgh

Photography and Data Management

The task is to accurately scan more than 250,000 photographic paper records of various quality and sizes, ensuring no damage to the originals and no loss of information in the copies. Initial considerations were to:

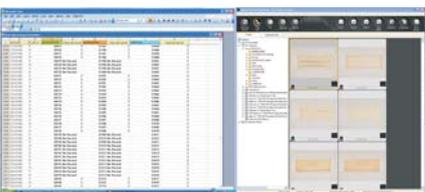
- establish the quality of reproduction required
- find the most cost-effective solution
- estimate the staff resources required
- determine the optimum technical details
- ensure rigorous quality control (QC)

After testing a number of options, a high resolution digital camera based system was chosen for the following reasons:

- it is able to resolve, both optically and digitally, the finest detail found on the records
- it is capable of capturing records up to eight times faster than a scanner based system
- it is less likely to damage the original records
- the auto-numbering feature of the camera greatly reduces the handling time for each record



The camera selected has a resolution of 21 megapixels and the lens used is a fixed focal length. It has a flat plane macro lens to minimise distortion, which is also checked by capturing an image of the background grid at the start and



end of a set of records. The camera is linked to and controlled by a PC (right). When digitised each batch of 1000 records requires 20Gb storage space. The raw image files in a camera specific format are converted to Jpeg 2000 (jp2) format and are archived.

1. magnetogram collections



2. metadata for each scan



3. path and filenames



A scan of an early record from Kew observatory is shown above.

The jp2 image collection is managed via a database which contains three main tables as shown. These are used as part of the QC process during scanning. Each paper record contains one or more recordings. There are two scans for each paper record: the front, containing the magnetic recording; and the back, which although usually blank, can occasionally contain important metadata.

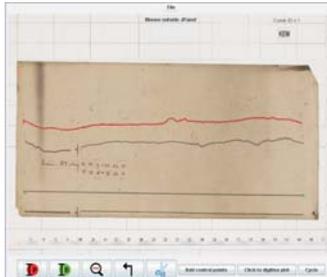
On-Line Delivery

All Kew magnetograms have now been digitised (74,132 scans, 37,066 plots). A web delivery system is currently under development that will enable access to these jp2 images. The system will be designed to enable fast browsing of the jp2 images at lower resolution as well as allowing access to the full resolution files if required.



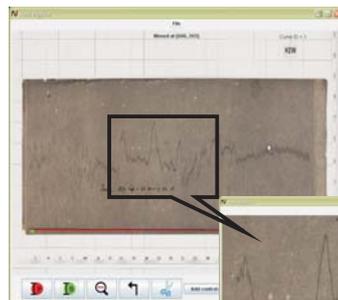
Digital Data Extraction Software

A Java programme is being developed by BGS to enable digital data to be extracted from the jp2 images. Version 1 of the program is not yet operational, however, we felt that it was appropriate for this poster to describe the work carried out to date and the future plans.



Once the baseline and variometer traces have been digitised, the data can be extracted. At present the values recorded are the perpendicular distances (in pixels) between the two lines. All output is saved in XML format. Later versions will have an add-on feature to allow the input of baseline and scale values and thus enable the magnetic field data to be extracted at the time of operation.

The screenshot below shows the program in use with a difficult example - a poorer quality image, with a lot of noise, on a magnetically active day.

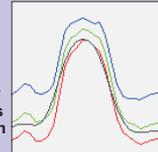


The plot far right demonstrates how the algorithm can fail and new control points are required.



Algorithm Details

The algorithm works by using control points that are selected by the user with mouse clicks. First, the user selects a start and end point on the baseline, to which a straight line is fitted according to the method of least squares. The user then selects the variometer trace, also with a start and end point. For both traces, the main digitising algorithm takes a vertical slice through the image of pixels, from the start point moving across the image, pixel by pixel. For each slice, a profile of the pixels is calculated and the distribution analysed to identify the peak in intensity (right). The pixel identified is then selected as the point on the required line.



For the variometer line the height of the slice is configured to avoid including auxiliary marks associated with the plots, such as pencil annotations, but large enough to include the line. Each slice extracted from the image is focussed over the trace by moving the slice up or down as it moves across the image. The magnitude of the change in the vertical direction is proportional to the distance from the centre of the slice to the peak in pixel intensity. This allows the slice to follow the gradient of the line. The pixel selected is recorded and its location plotted onto the image (in red) so that the predicted trace can be displayed to the user.

In the event of failure, there is a 'cut' function and a tool for adding control points to accurately guide the digitisation of more complex graphs. There is also a zoom function to allow the user to add control points more precisely.

Future Plans

The programme of work to complete the magnetogram scanning project is likely to continue for many years. The robust method adopted will ensure that all user requirements are met now, and in the future.

Following completion of the web delivery system, online public access to the Kew magnetograms will be released and the web site publicised to the community.

The Greenwich magnetogram collection will be the next set to be scanned. Currently, the camera and other settings are being optimised for this set, with work due to start in September. An additional complexity with the early Greenwich magnetograms is that their bindings must be removed before they are scanned.

Plans are also being made to digitise the complete set of yearbook records for the UK observatories.

The current version of the digital data extraction software is very near to being ready for use. Software development will continue, both to improve the recognition accuracy and to make operation of the software fast and easy for the user - essential, given the amount of data that needs to be digitised.