NATURAL ENVIRONMENT RESEARCH COUNCIL

Predicting Extremes in European Geomagnetic Activity

Ewan Dawson (ewan@bgs.ac.uk), Sarah Reay, Alan Thomson
British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom

Abstract

Rapid variations in the geomagnetic field constitute a natural hazard, e.g. for navigation and to power grids and pipeline networks. In order to better allocate resources towards mitigating these risks, we must be able to model the recurrence of extremes of geomagnetic activity over many years.

However, the data we have from which to develop such a model, in the form of continuous series of 1-minute samples of the geomagnetic field, typically stretch back less than 40 years. Without a longer record, it is difficult to construct a clear picture of the magnitude and frequency of extremes in geomagnetic activity.

We therefore apply the statistical technique of 'extreme value analysis' on a number of decades of geomagnetic data recorded at observatories across Europe, and in doing so arrive at an estimate of the maximum field strength and time-variation that might be observed once in every 100 and 200 years.

Extreme Value Theory

We use a Generalised Pareto Distribution (GPD) to describe the tail of the distribution of geomagnetic activity (see e.g. Coles, 2004). The GPD is a unification of the Gumbel, Frechet and Weibul distributions, widely used in the scientific literature when modelling extremes in variables. By fitting a GPD curve to the extreme values in our data set, we can make estimates about the probability of geomagnetic conditions more extreme than those in our data set, and thus provide estimates of the largest extremes likely to be observed over a given period.

To accurately characterise the probability of extreme values, we must exclude all non-extreme samples from our data. This can be achieved using a 'point over threshold' approach, where some value is chosen as a threshold for extreme geomagnetic activity.

There are some assumptions implicit in the theory behind the GPD which must be examined: namely that the data are stationary (show no time-dependancy) and the probablility of a particular sample exceeding the threshold is not dependant on the value of previous samples; that is, the data are independent.

Clearly, geomagnetic data are not independent; storms lasting several hours are likely to produce a number of extreme values, which cannot be considered to be independent of one-another. This may be dealt with through de-clustering the data; this technique is described later.

Neither are the data stationary; the geomagnetic field exhibits cyclic behaviour with many periods from days to years. However, we assume there to be no significant time-dependance on the scale of centuries and that any non-stationarity does not affect the results at these time scales.

Data Preparation

One-minute geomagnetic time series of *H* (horizontal field) and *D* (declination) were downloaded from the World Data Centre for Geomagnetism in Edinburgh (wdc.bgs.ac.uk) for 29 European observatories.

The observatories were chosen to provide a representative spread of locations across the continent, covering a range of magnetic latitudes and for which there exists continuous data spanning a number of years.

From *D* and *H*, we computed time series of the variations from quiet levels due to the influence of the external field. These were computed for each observatory by removing the quiet mean level, which was established for each month from the five 'International Quiet Days' as determined by the International Service of Geomagnetic Indices.

A further two time series, *dD/dt* and *dH/dt*, were constructed by computing the difference between successive sample in the original *D* and *H* time series. (Fig. 1)

Our analysis was performed on the four time series for each of the 29 observatories.

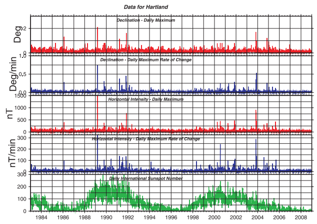


Figure 1. *H* (nT), *D* (degrees) residuals and one-minute rates-of-change for the Hartland mid-latitude observatory (one minute data, 1983-2009). Daily maxima of the absolute residuals are shown, as is sunspot number, to identify any solar cycle dependence.

Extreme Value Analysis

Threshold Selection

Before fitting a GPD curve to the data, we first define the threshold which determines which values are to be considered extreme. All data points below this threshold are discarded.

An appropriate threshold for each of the variables was determined by plotting the scale and shape parameters of the resulting GPD for a range of thresholds. The ideal threshold should be low enough to allow for a meaningful number of samples, but high enough that the modified scale parameter is constant and the shape parameter linear (within error-margins), above the chosen threshold . (Fig 2).

We found that setting the threshold at the 98.6th percentile was reasonable for each variable at most observatories.

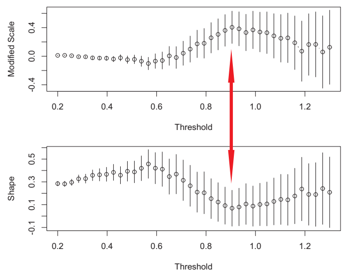


Figure 2. Response of modified scale and shape parameters of GPD function to threshold. The red arrows indicate a reasonable choice of threshold for these data (residual D values at Eskdalemuir observatory).

De-clustering

Clusters of extreme values occur during geomagnetic storms. This results in statistical dependency in the data, which must be eliminated to meet the assumptions of the model. We identified clusters by looking for extreme values that were not separated by at least one day. Only the peak value from each cluster was retained.

Determine GPD parameters

Applying the threshold and de-clustering reduces the data set by more than 99.9%. The remaining data are fitted to a GPD using the 'R' statistics software, along with the 'eXtremes' package. The model can be plotted as a 'return-level plot' (Fig 3).

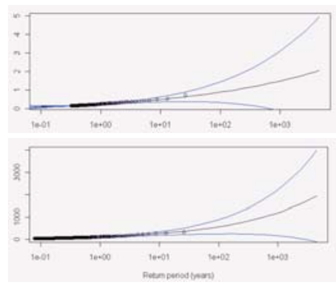
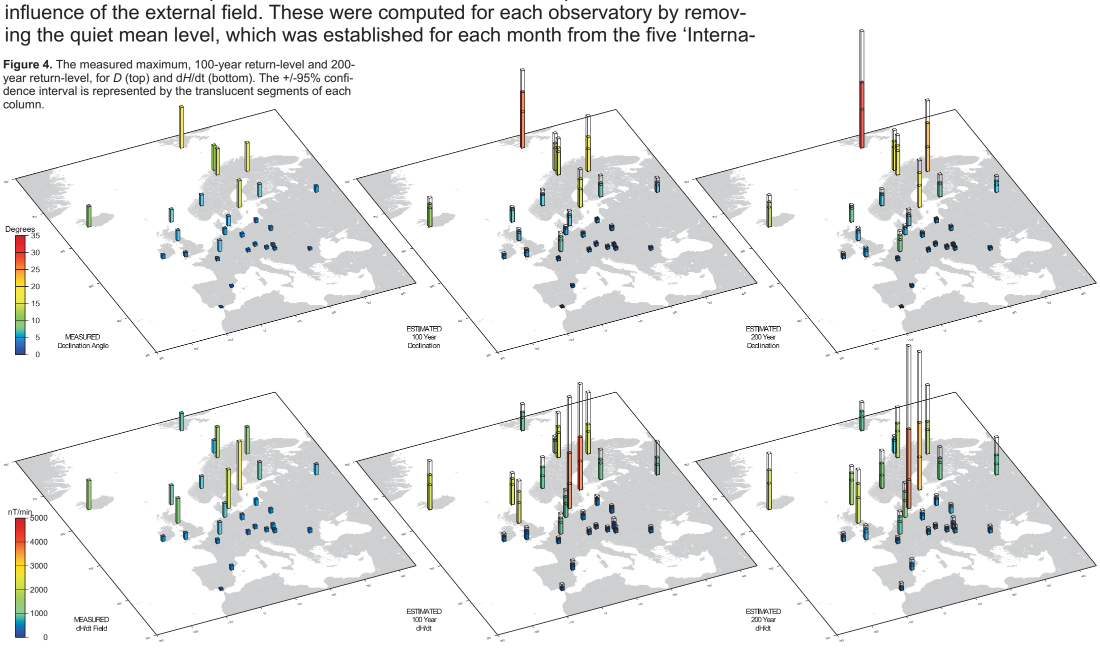


Figure 3. Return periods for observed Hartland *D* (upper) and *H* (lower) residuals (circles) and the fitted and extrapolated GPD (line) to each of *D* and *H*. Vertical scales are degrees and nT respectively; horizontal scale is time in years. The blue lines are the approximate symmetric +/-95% confidence limits from the fit of model to data, via an *eXtremes* function.

Figure 4. The measured maximum, 100-year return-level and 200-year return-level, for *D* (top) and d*H*/dt (bottom). The +/-95% confidence interval is represented by the translucent segments of each column.



Results

For each observatory we have computed the peak variation and rate-of-change predicted by the GPD to occur over periods of 100 and 200 years. The results for two of the four time series are summarised in Figure 4. The results from the UK observatories at Lerwick, Eskdalemuir and Hartland are shown below.

	D (residual)	dD/dt	H (residual)	dH/dt
LER	**100 year: 6.64** (5.76, 7.51) **200 year: 7.59** (6.57, 8.62)	**100 year: 2.90** (2.30, 3.50) **200 year: 3.32** (2.63, 4.02)	**100 year: 3441** (3133, 3749) **200 year: 3808** (3462, 4154)	**100 year: 1297** (996, 1599) **200 year: 1582** (1208, 1956)
ESK	**100 year: 4.53** (3.57, 5.78) **200 year: 5.03** (3.84, 6.46)	**100 year: 3.42** (1.86, 5.53) **200 year: 4.44** (2.13, 7.28)	**100 year: 4130** (3115, 5144) **200 year: 4855** (3641, 6069)	**100 year: 1591** (1016, 2166) **200 year: 1995** (1261, 2729)
HAD	**100 year: 3.08** (2.29, 3.86) **200 year: 3.59** (2.62, 4.52)	**100 year: 0.73** (0.58, 1.08) **200 year: 0.80** (0.62, 1.19)	**100 year: 2164** (1656, 2673) **200 year: 2652** (2016, 3288)	**100 year: 434** (294, 600) **200 year: 510** (323, 710)

Future Work

This statistical analysis could be improved by, for example, treating variables D (H) and dD/dt (dH/dt) as components of the same multi-variate statistic.

It should also be possible to extend the application of this technique to look at extremes in global activity levels, dependant on data availability.

Acknowledgments

We would like to acknowledge scientific institutes in Europe for providing their magnetic data via INTERMAGNET and the World Data Centres. Colleagues at BGS are also thanked for comments on this work.

References

Coles, S., 2004 An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag, London, ISBN 1-852334-59-2.

Gilleland, E and Katz, R. W., 2005. Tutorial for the Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics, http://www.assessment.ucar.edu/toolkit.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, www.R-project.org.

RAS National Astronomy Meeting
Glasgow, Scotland April 2010
Session: MIST Science Session P18

RAS NAM2010
with the UK Solar Physics & MIST meetings
12-16 April 2010, University of Glasgow, UK