Magnetic Observatory Data Products for Space Weather Applications

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

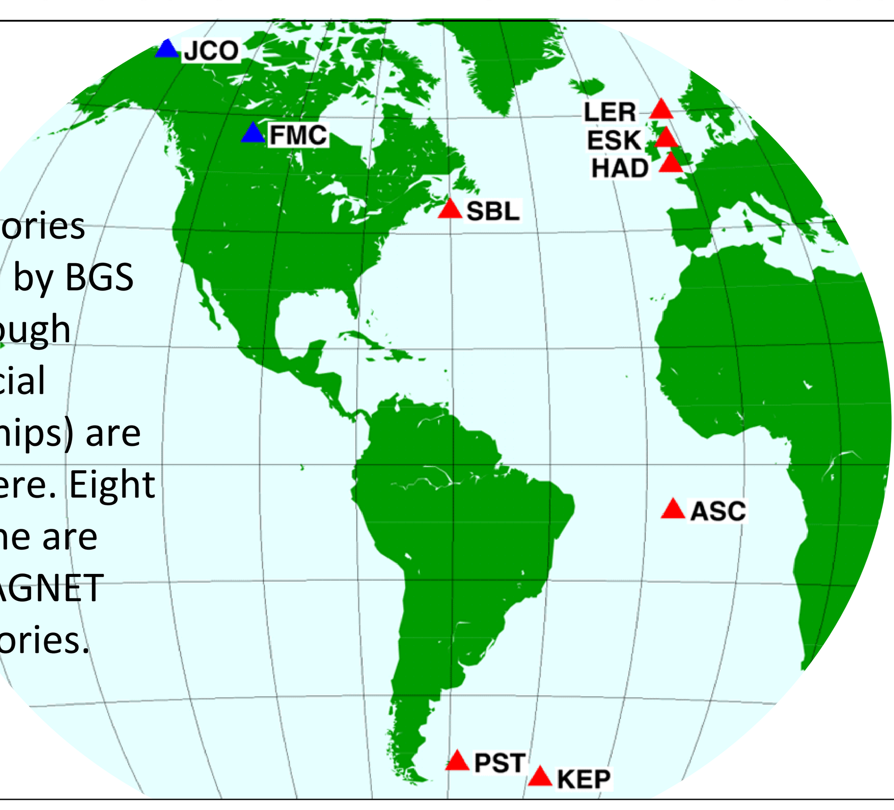
Ellen Clarke (ecla@bgs.ac.uk), Gemma Richardson, Alan Thomson, Orsi Baillie, Sarah Reay, Thomas Humphries, John Williamson and Laurence Billingham*
British Geological Survey, Edinburgh, UK; (*Now at Mercurytide, Edinburgh, UK)

1. Introduction

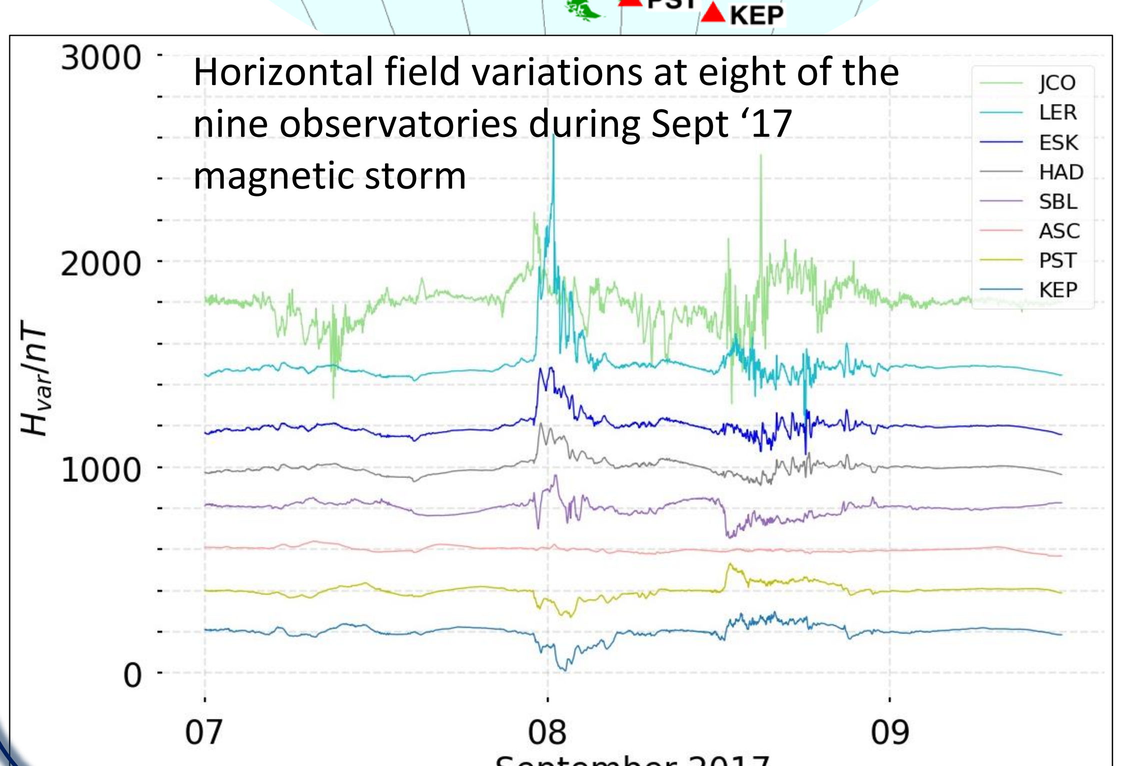
Space weather is on the UK government's national risk register. Magnetic observatory measurements provide the underlying capability for real-time dissemination of information and advice on geomagnetic activity and space weather hazard. Long-term operation of observatories enables continuous monitoring of activity levels and is therefore a key component. Operational products derived from the observatories include real time local activity (Hourly Standard Deviation or HSD, d**B**/dt, K, K_{BI}, A_{BI}, DRX) and estimated global indices (Kp_{est}, ap/Ap_{est} and aa/Aa_{est}) updated at 5 minute intervals. Monitoring is supplemented with forecasts of local and global 3-hourly and daily activity indices from automated algorithms (ARIMA and Neural Net) as well as human-derived one-, two- and three-day-ahead categorical activity forecasts on a daily basis.

2. Geomagnetic Observatory Network: Data and Indices

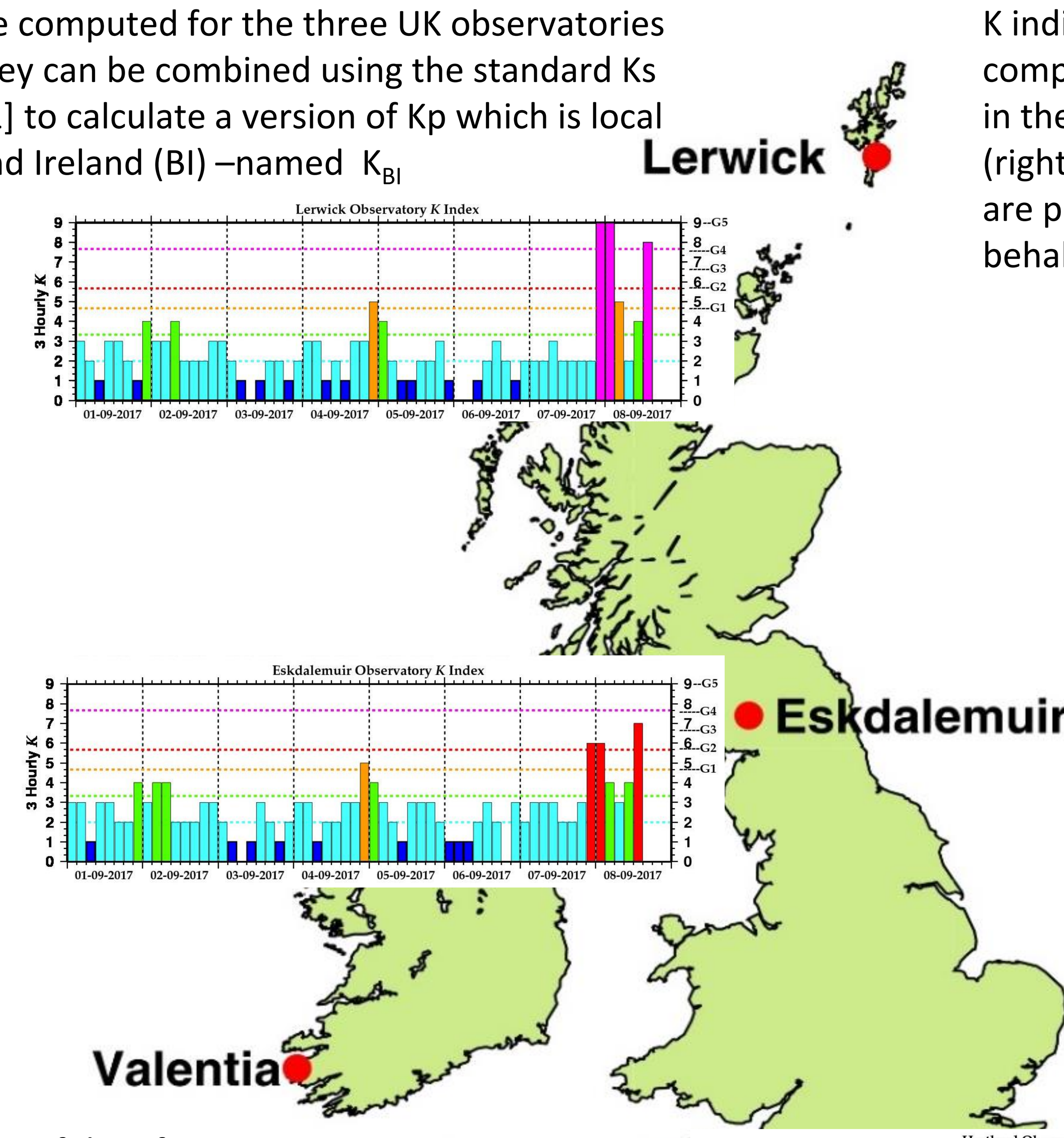
Accurate, timely and reliable space weather products rely on high quality, accurate and reliable observatory data. The real-time processing of data from the BGS observatory network and other INTERMAGNET-standard observatories provides the primary essential ingredients for the derivation of indices and forecasts described here.



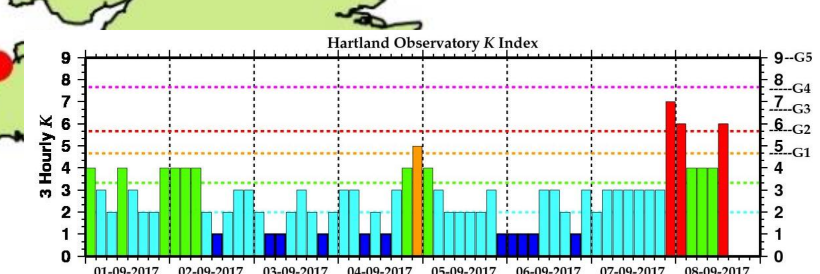
Observatories operated by BGS (two through commercial partnerships) are shown here. Eight of the nine are INTERMAGNET observatories.



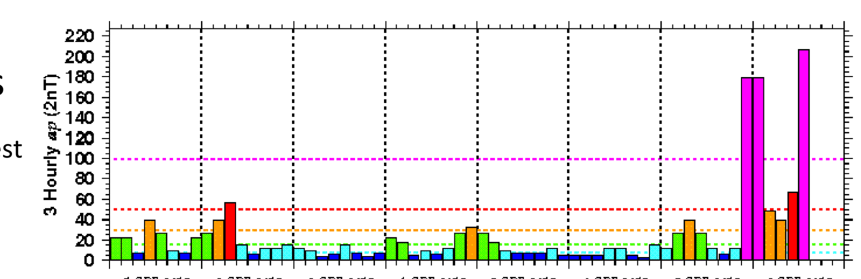
K indices are computed for the three UK observatories (below). They can be combined using the standard Ks algorithm [1] to calculate a version of Kp which is local to Britain and Ireland (BI) –named K_{BI}

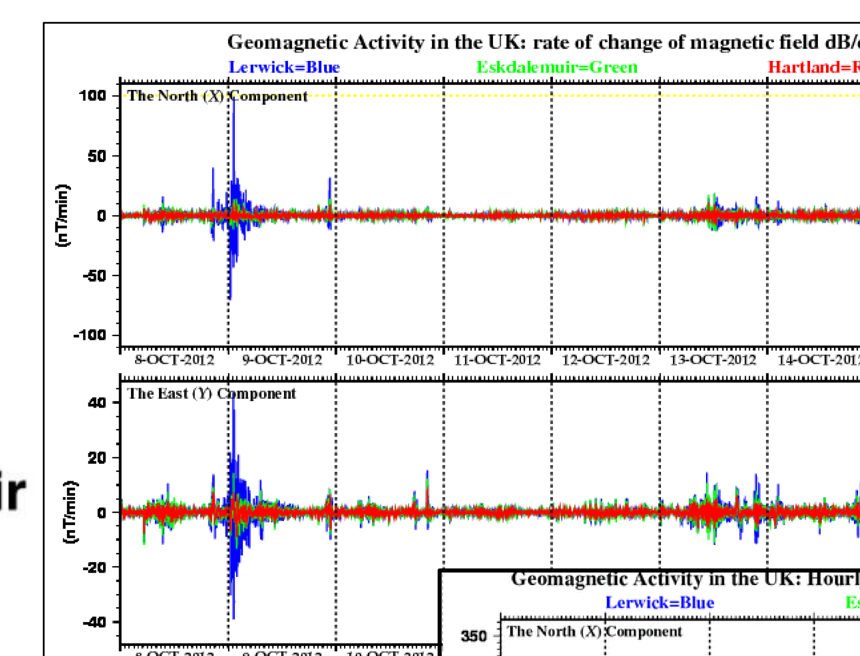


Inclusion of data from VAL operated by Met Eireann along with data from Northern Ireland in the derivation of K_{BI} is planned in the future.

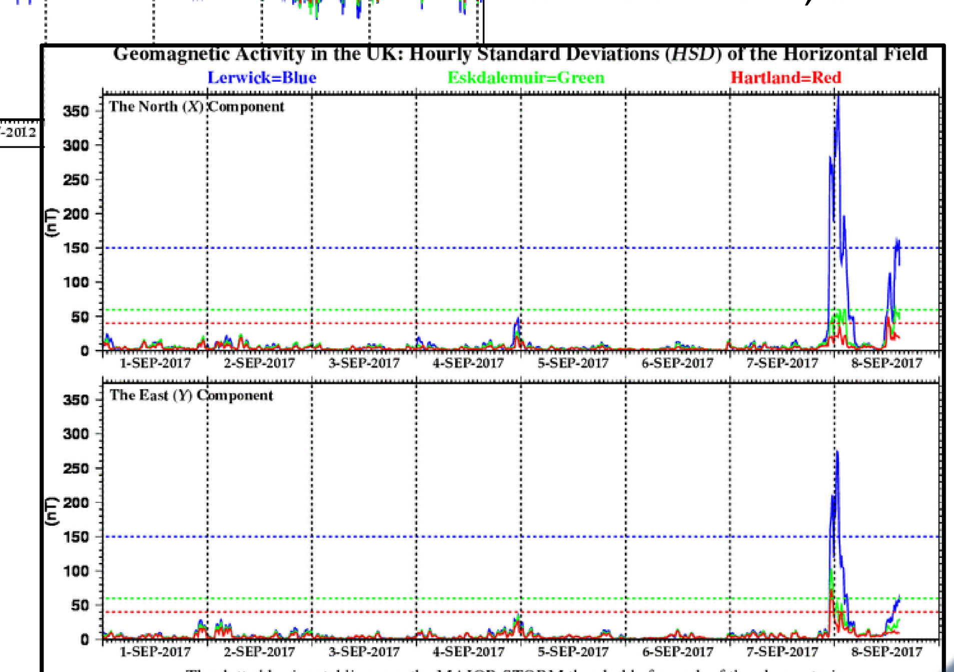


K indices are also collected or computed for other observatories in the Kp network to calculate ap_{est} (right). Note: definitive Kp and Ap are provided by GFZ, Potsdam on behalf of ISGI [2].

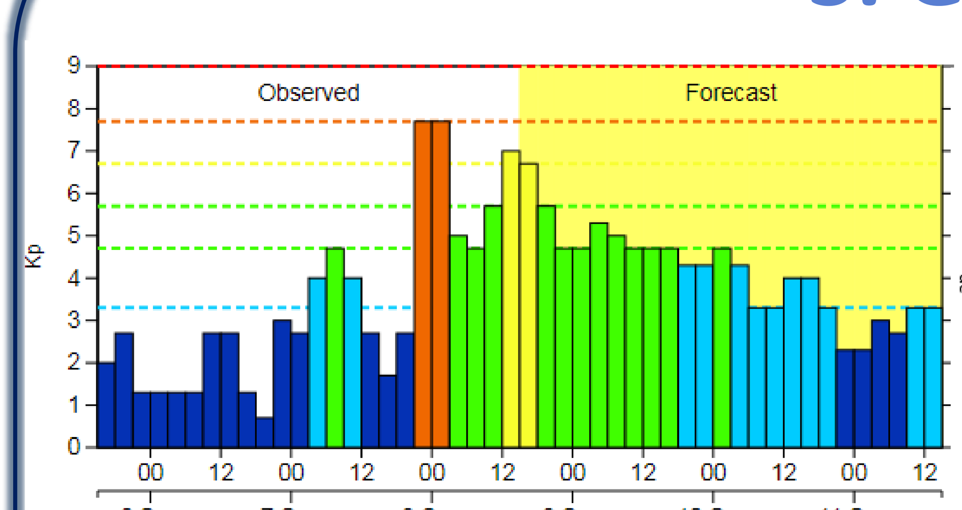




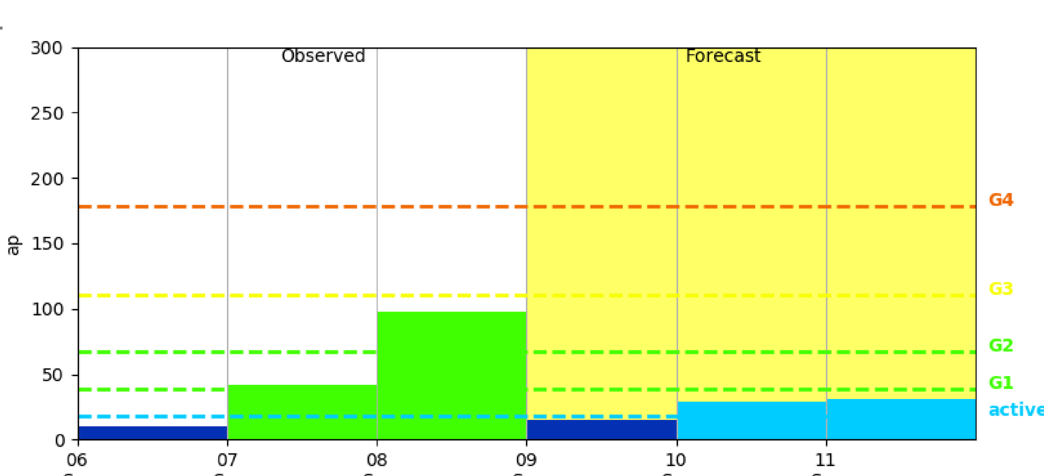
HSD (below) and d**B**/dt (left) at the UK observatories are provided in near real-time. These products form part of the MAGIC (Monitoring and Analysis of GIC) service provided to National Grid, UK.



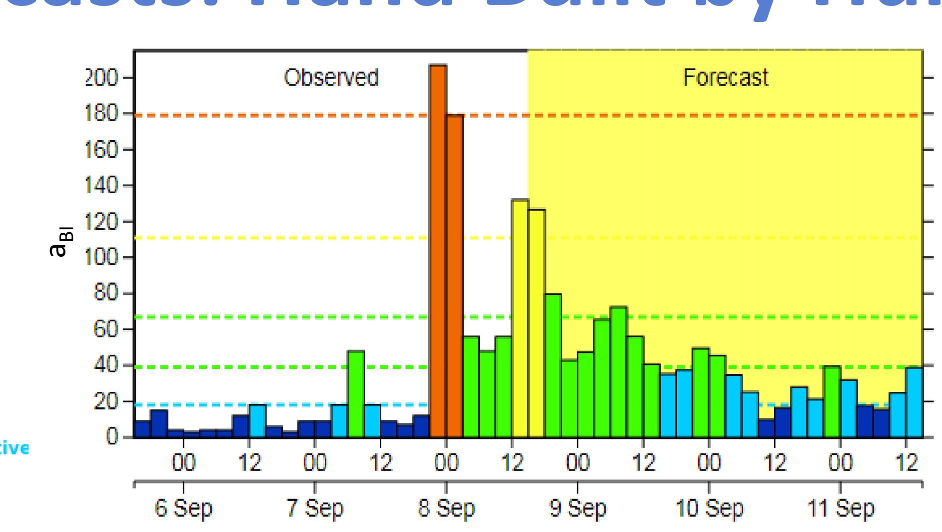
3. Geomagnetic Activity Forecasts: Hand Built by Humans

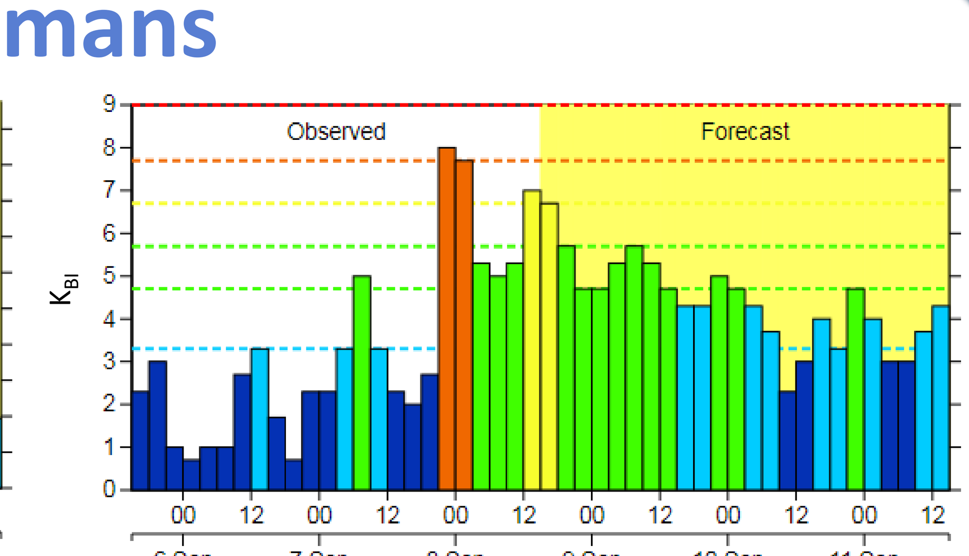


Real-time predictions of Kp_{est} using ARIMA method [3]. Used operationally for 27-day predictions at the European Space Operations Centre since 1992, it was further developed for 3-hourly values in 2015.



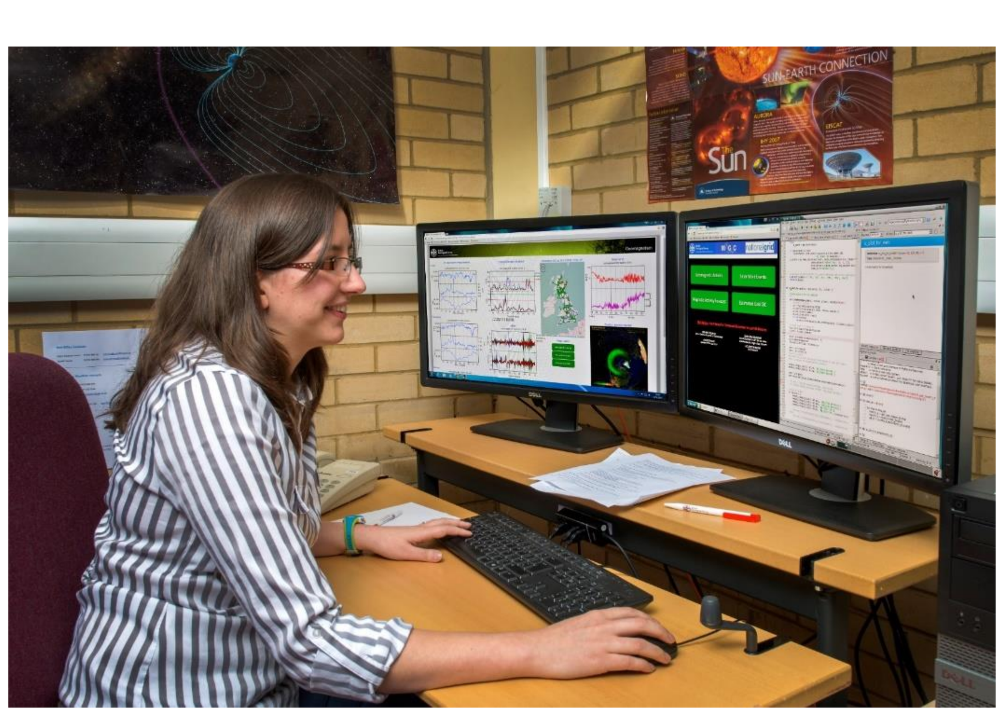
Real-time observed and predictions of daily Ap_{est} using a Neural Network method [4].





Real-time observed and predictions of local 3-hourly a_{BI} and K_{BI} using the ARIMA method

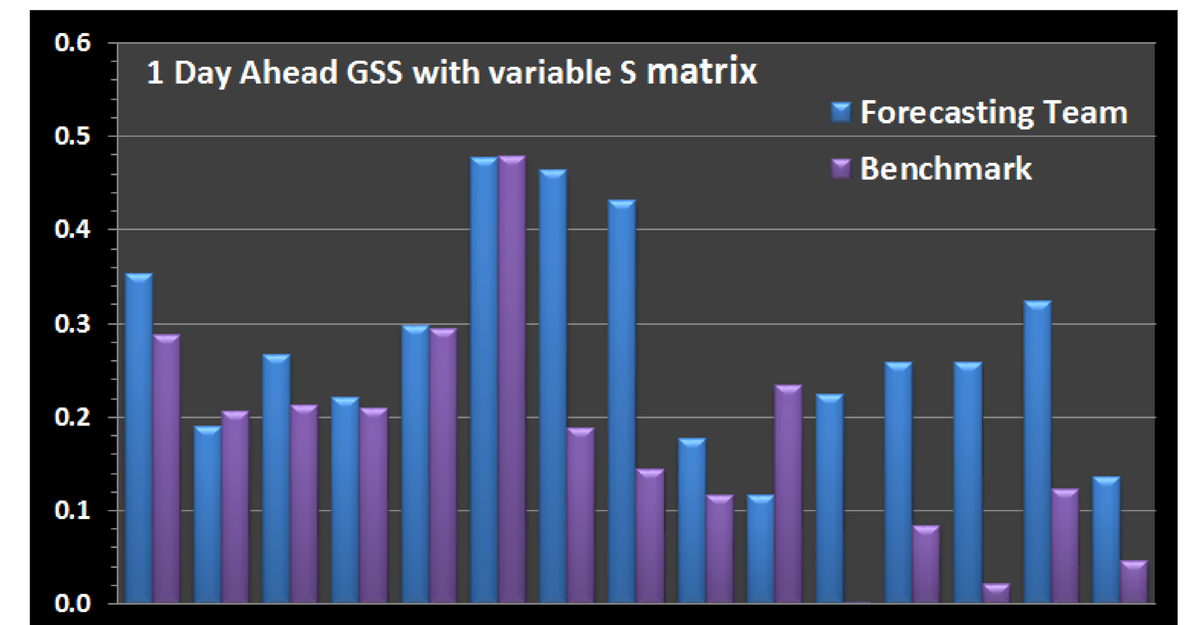
- Local indices are more appropriate than planetary for most ground-based applications
- Although spanning 3-hourly time windows, the real-time operational cadence of all these products is 5 minutes



A forecaster (above) making 1, 2 and 3-day ahead forecasts. Various solar and solar wind observations, data and models available in the public domain, as well as in-house products are analysed and interpreted.

Below is the forecast made by the duty forecaster at 10.38UT on 7th Sep 2017. The storm, which was predicted due to a CME associated with the ~X10 solar flare of the 6th Sep, peaked at G4 and averaged G2. (storm shown above and in section 2). The human forecast on this occasion, although not perfect, out-performed any automated algorithms (ARIMA, Neural Net) running at BGS at the time.

Forecast period (noon-to-noon GMT)	Forecast Global Activity level	
	Average	Max
7 SEP-8 SEP	ACTIVE	STORM G3
8 SEP-9 SEP	STORM G2	STORM G3
9 SEP-10 SEP	ACTIVE	STORM G2



Skill of human forecasters over 13 years against a benchmark of persistence and recurrence. A higher score indicates more skill. The S matrix of the Equitable (Gerrity) Skill Score (GSS) [5,6] was adapted to account for geomagnetic activity levels for each year[7]. Both the individual example (left) and the long term skill scores (above) are evidence of the need for human interpretation in geomagnetic activity forecasting.

4. Finding the 'Right' Geomagnetic Parameter for the Job

Kp (and the related Ap) are used extensively in space weather applications and models. Kp has become an (informal) standard for geomagnetic activity, despite its well documented limitations in both space and time. There are many applications where these drawbacks are not significant and its use is reasonable. However, there are some applications where it could be misleading to consider only Kp, in particular where localised geomagnetic activity levels are required and over a shorter time span than 3 hours. One such example is the monitoring of geomagnetically induced currents (GIC) in power systems.

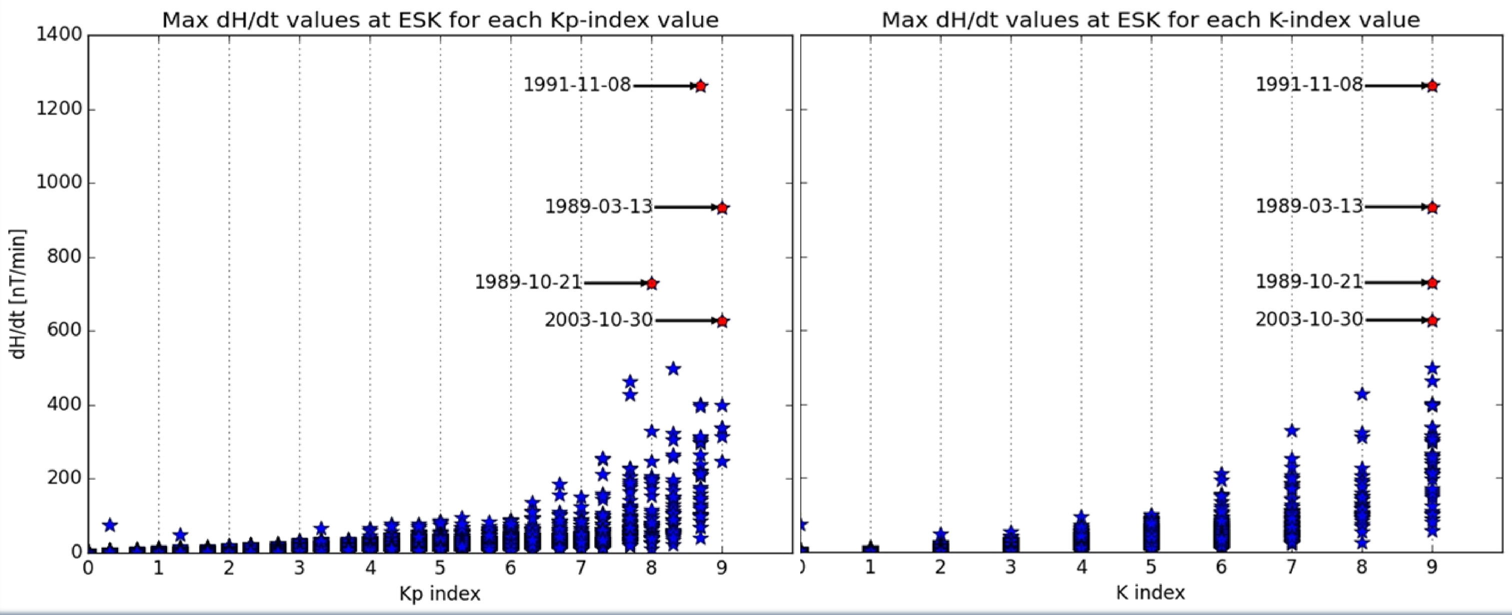
Kp	BGS categories since 2014		NOAA G-scales	
	Category	Description	Category	Description
<3+	QUIET	Kp < 3+		
3+	ACTIVE	3+ < Kp < 5-		
4-				
4o				
4+				
5-	STORM G1	5- < Kp < 5+	G1	Kp = 5
5o				
5+				
6-	STORM G2	6- < Kp < 6+	G2	Kp = 6
6o				
6+				
7-	STORM G3	7- < Kp < 7+	G3	Kp = 7
7o				
7+				
8-	STORM G4	8- < Kp < 9-	G4	Kp = 8
8o				
8+				
9-				
9o	STORM G5	Kp = 9o	G5	Kp = 9

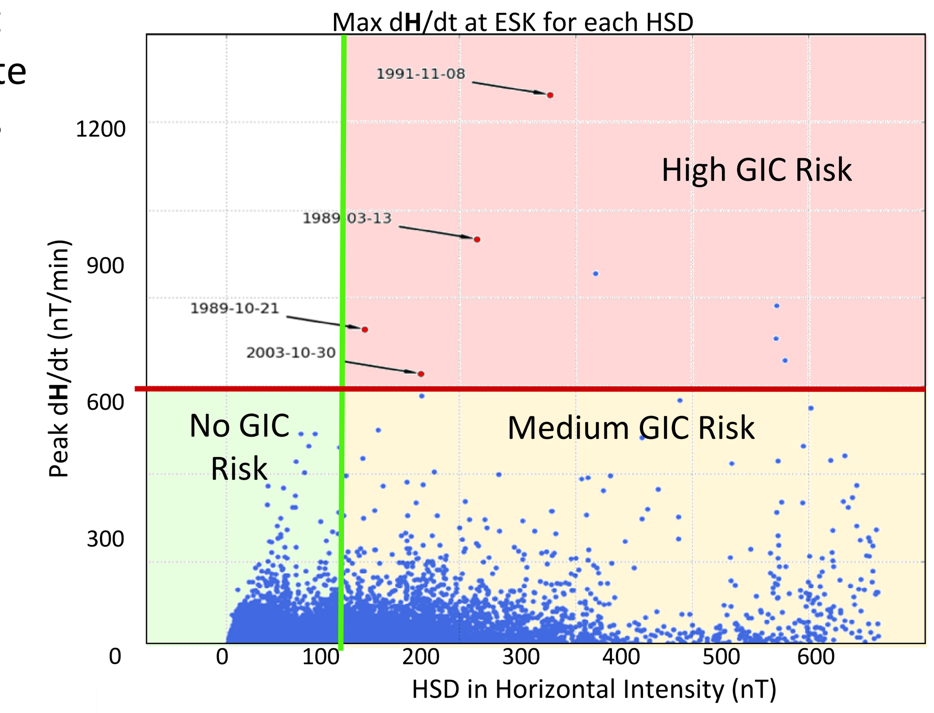
Kp categories of geomagnetic activity map to the NOAA G scales. **Power companies in the UK are only likely to take action at G5 storms – therefore further category differentiation is required for extreme events.**

A parameter that will represent likely GIC flowing in the system and one that can be forecast to provide warnings is required. GIC is known to relate to rate of change of the magnetic field (d**B**/dt). Peak d**H**/dt at ESK compared to Kp is shown (below left). The red starred events indicate storms with known GIC impacts. d**H**/dt is widely spread at high levels of Kp and we know that Kp>8 doesn't necessarily result in problems.

Replacing Kp with local K (below right) shows the known impact events were all for the highest K and highest d**H**/dt. However, many K=9 periods do not correspond to high d**H**/dt.

Conclusions at this stage are that a monitor for GIC needs to be at a higher time resolution and local to the power grid location. Planetary indices are therefore of limited use and could be misleading.

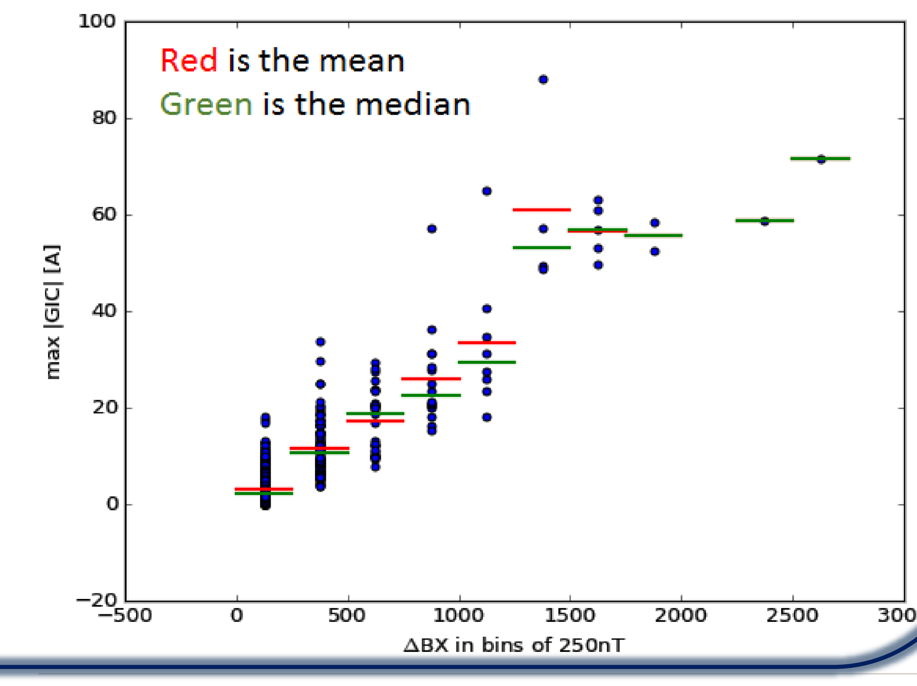




HSD (shown in section 2 above), a parameter that is physically related to GIC [8], is an improvement. Computed with a 5-minute cadence, HSD has been used operationally to monitor GIC in the UK since 2000 [9].

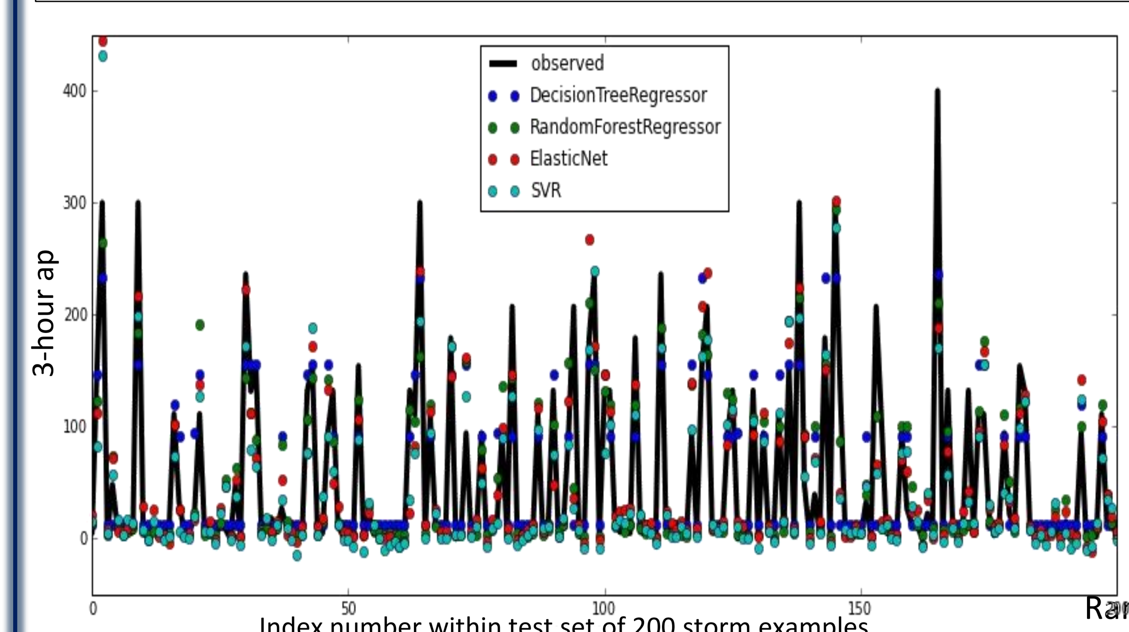
One option for identifying a threshold for warnings/alarms could be a combined approach as shown diagrammatically above.

Further work to determine the most suitable local parameter(s) for operational GIC monitoring is on going - exploiting the correlation between modelled GIC (accounting for both Earth conductivity and power network – see presentation on Friday by Richardson in Session 15 for more on GIC modelling) and various field parameters in both the time and frequency domains [10]. One example is shown below. These studies are indicating a higher linear correlation between GIC with variations in **B** than with d**B**/dt, although it is too early to draw conclusions from this work in progress.



5. Activity predictions with Machine Learning (ML) – Better or Not?

The algorithm for predicting 3-hour ap using the ARIMA method relies entirely on patterns within the time-series itself. It is now well established that making use of precursor data - solar and solar wind - can improve geomagnetic predictions. Exactly the best way of making use of the available data in an operational sense, is not yet established. In 2014 [11] work was carried out to evaluate Machine Learning (ML) algorithms [12] in predicting ap at one 3-hour period ahead, using the ap time series and ACE solar wind measurements as input. Further work in 2016 [13] found that the method was more useful for activity categories, such as NOAA's G storm scale. Selected results are highlighted here.



Evaluation of various methods against observed ap [11]

(with 1000 samples)	rms	% within ± 5	% within ± 10	HitRate	HSS	FAR
DecisionTreeRegressor	41.40	16.42	52.24	0.89	0.84	0.11
RandomForestRegressor	34.26	56.22	69.65	0.97	0.89	0.12
ElasticNet	36.48	34.33	56.72	0.95	0.91	0.08
SVR	38.18	35.82	63.68	0.90	0.88	0.07
LinearRegression	37.26	34.83	54.73	0.95	0.91	0.08
ARIMA	49.04	59.79	70.10	0.82	0.84	0.04

Performance (left) for the best of the ML algorithms showed:

- all ML methods out-perform ARIMA in Skill (HSS), Hit Rate and rms
- ARIMA more accurate (within ±5/±10 ap units) and has the best False Alarm Rate (FAR)
- best all rounder (in this case) is `Random Forest Regressor' although it does have the poorest FAR

(n=3625)	HSS	FAR
Forecasters 1day	0.34	0.54
BM 1day	0.29	0.64
Forecasters 2days	0.27	0.59
BM 2days	0.18	0.76
Forecasters 3days	0.11	0.00
BM 3days	0.14	0.80

Similarly the performance in terms of HSS and FAR (above) of human forecasters at BGS over 13 years were measured against a benchmark (BM) of persistence and recurrence [7]. These are categorical forecasts of storms at 1, 2 and 3, days ahead. The scores above are not directly comparable to the equivalent for the ML 3-hourly predictions. Further work to establish a true like for like comparison is needed. Note that the FAR=0.0 for 3-day ahead is due to there being no predictions made.

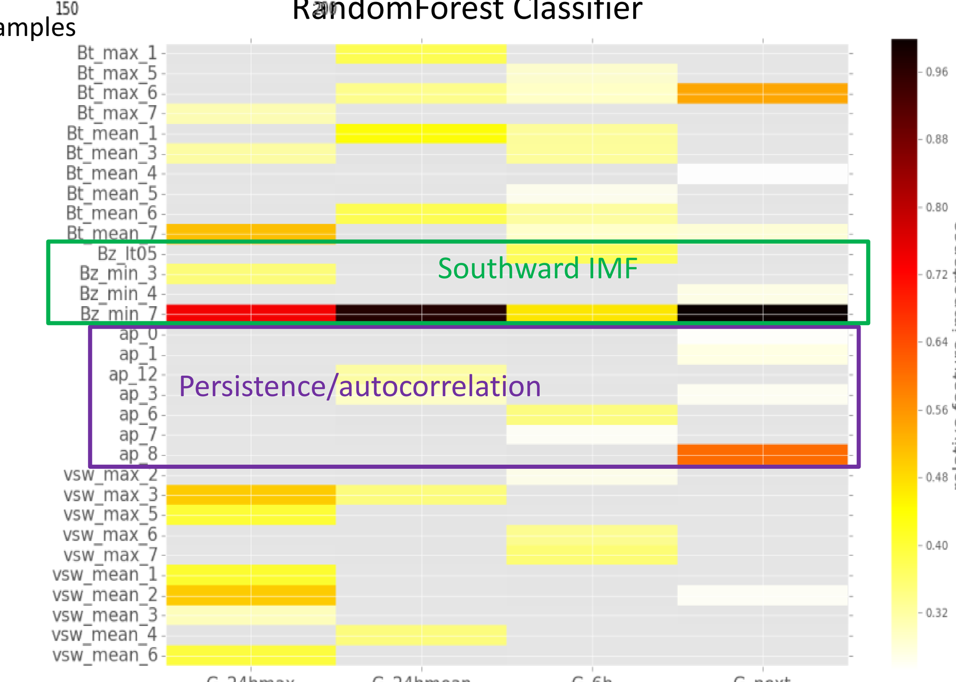
This work (2016) [13] showed:

→advantages to using classifications of activity rather than numerical values

→ML methods allow the discovery of the most important features - less 'black box' than Neural Networks

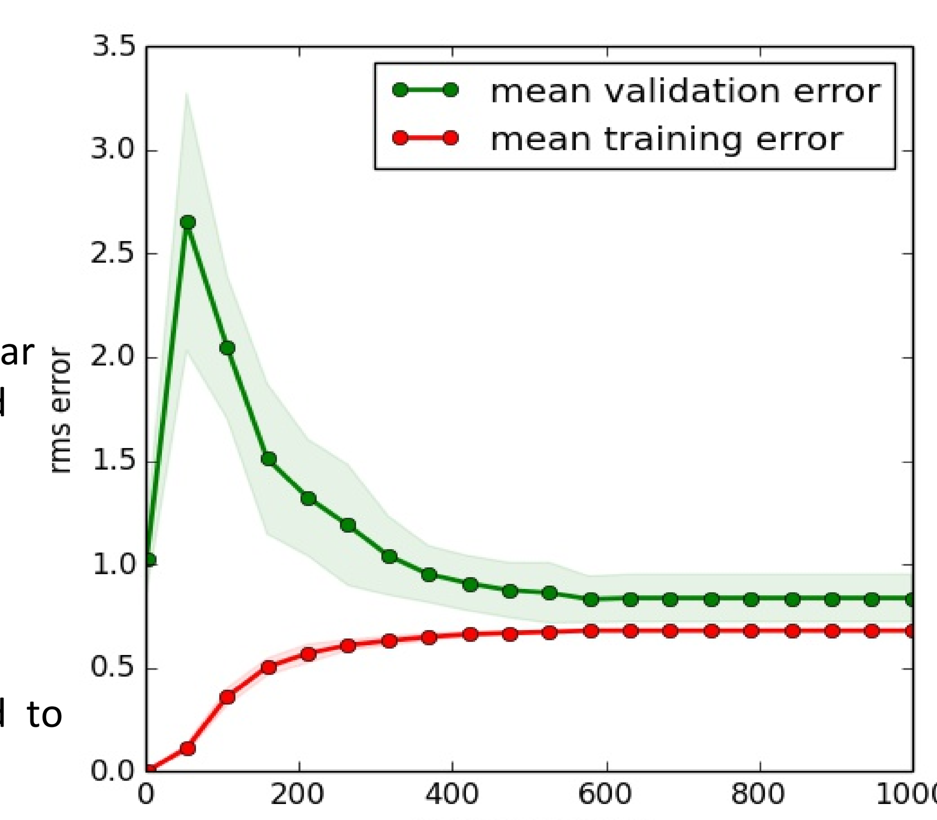
→The most informative features (shown right) for the best method found so far - the 'RandomForest Classifier'.

RandomForest Classifier



Training converges quickly (right) – this will enable the inclusion of more features in future developments e.g. solar data, radio bursts, solar wind coupling functions

Work is now required to advance the ML code to an operational level. Further performance tests compared to other operational forecasts would then be possible.



Summary and Future Work

The use of geomagnetic data and indices for BGS space weather applications has been reviewed and a summary has been given on the research carried out to enable and improve on present-day operational capabilities.

Forecasting activities have been examined, and we argue that the inclusion of a human forecaster is likely to provide more useful forecasts than entirely automated computer-based methods. Despite this, development of algorithms, including ML, will continue to try to improve the accuracy of the operational automated predictions being made. Integration of an on-going forecast evaluation would also be a useful addition.

Work to establish the best monitoring parameter for GIC is continuing, which, as well as being useful for power companies, will feed directly into a new project recently started in the UK to cover *Space Weather Impact on Ground-based Systems* (SWIGS) – see poster later this week (session 15). A network of new magnetometer station pairs across the UK is planned to measure and assess GIC in power, pipeline and railway networks. These new data sets will complement those from the existing long-running magnetic observatories in Britain and Ireland and collectively provide an invaluable resource towards the on-going space weather research activities, some of which have been described here.

References

[1] Bartels, J. (1949) - The standardized index Ks and the planetary index Kp, *IATME Bulletin 12b*, **97**.

[2] ISGI – The International Service for Geomagnetic Indices. www.isgi.unistra.fr

[3] Thomson, A.W.P., Clark, T.D.G. and Kerridge, D.J. (1992) Computer Algorithms and FORTRAN Programs for Forecasting Solar and Geomagnetic Activity in the Short-term, *British Geological Survey Technical Report* WM/92/19C

[4] Thomson, A.W.P. (1996), Non-linear predictions of Ap by activity class and numerical value. *Pure and Applied Geophysics*, Volume **146**, Number 1, Page 163

[5] Gandin, L. and Murphy, A.H. (1992) Equitable skill scores for categorical forecasts. *Mon. Wea. Rev.*, 120, 361-370.

[6] Gerrity, J.P.Jnr. (1992) A Note on Gandin and Murphy's Equitable Skill Score. *Mon. Wea. Rev.*, 120, 2709-2712.

[7] Clarke, E. and Thomson, A.W.P. Forecast Evaluation as Applied to Geomagnetic Activity Categories (2013), Conference Presentation, European Space Weather Week 10. www.stce.be/esww10/contributions/public/talks/Session12/05-ClarkeEllen/Clarke_ForecastVerification.pdf)

[8] Beamish, D., Clark, T.D.G., Clarke, E., and Thomson, A.W.P. (2002) Geomagnetically induced currents in the UK: geomagnetic variations and surface electric fields, Journal of *Atmospheric and Solar-Terrestrial Physics* **64**, 1779 – 1792

[9] Thomson, A.W.P., Clark, T.D.G., Clarke, E. and McKay, A. (2002) Monitoring and Forecasting of Geomagnetic Disturbances in the British Isles in 2000-2001. *British Geological Survey Internal Report*, CR/02/75.

[10] Marshall, R.A., E. A. Smith, M. J. Francis, C. L. Waters, and M. D. Sciffer (2011), A preliminary risk assessment of the Australian region power network to space weather, *Space Weather*, **9**, S10004, doi:10.1029/2011SW000685

[11] Billingham, L. and Kelly, G. (2014) Improving Operational Geomagnetic Index Forecasting. Conference Presentation, European Space Weather Week 11. www.stce.be/esww11/contributions/public/Session13/S13-HPE-04-BillinghamL/

[12] Hastie et al., The Elements of Statistical Learning Data Mining, Inference, and Prediction, Springer 2009(II)

[13] Billingham, L. and Kelly, G. (2016) An application of machine learning to geomagnetic index prediction. Conference Presentation, European Space Weather Week 13. www.stce.be/esww13/contributions/public/S11-O1/S11-O1-13-BillinghamLaurence/